DOCUMENT RESUME

ED 108 981 SO 008 263

AUTHOR	Jones, Frank Geoffrey
TITLE	The Effects of Mastery and Aptitude on Learning, Retention, and Time.
PUB DATE	74
NOTE	166p.; Ed.D. Dissertation, University of Georgia
AVAILABLE FROM	Geography Curriculum Project, Room 107, Dudley Hall, University of Georgia, Athens, Georgia 30602 ($5.00)
EDRS PRICE	MF-$0.76 HC-$8.24 PLUS POSTAGE
DESCRIPTORS	Academic Achievement; *Aptitude; Discrimination Learning; Educational Experience; Educational Research; Geography; *Geography Instruction; Grade 7; Individual Differences; Junior High Schools; *Learning; Learning Theories; Literature Reviews; *Programed Instruction; Retention; *Teaching Methods; Time Factors (Learning)
IDENTIFIERS	Geography Curriculum Project; *Mastery Learning

ABSTRACT

The major purpose of this study was to compare self-instructional mastery and nonmastery treatments to determine if there are differences in learning, retention, and time-to-testing of high, middle, and low aptitude students. Twenty grade 7 classes from the Savannah-Chatham County School District served as the experimental population. Students were tested for placement in one of three levels of aptitude; then, classes were randomly assigned to two groups and treatment was randomly assigned to groups. The nonmastery treatment received a student text and a workbook which contained prescribed activities and a single review test for each chapter. The mastery treatment received the same student text; however, the chapters in the workbook contained two review tests. If the criterion level was not attained in the first review test, mastery students were required to correct and relearn material and then take a second review test. A multiple choice test and recall test was administered to measure learning and retention of the content materials. Findings showed that differences in aptitude were not reduced when self-instructional materials were used. An implication of this study is, however, that the lack of teacher monitoring in administering the review tests may have contributed to the poor performance of low aptitude students, since typically low aptitude students require close supervision. (Author/ND)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED108981

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

THE EFFECTS OF MASTERY AND APTITUDE ON LEARNING, RETENTION, AND TIME

by

FRANK GEOFFREY JONES

Dip. P.Ed., Mercer House, 1962
B.Ed., University of Calgary, 1970
Dip. Ed., University of Calgary, 1971
M.Ed., University of Calgary, 1972

A Dissertation Submitted to the Graduate Faculty
of the University of Georgia in Partial Fulfillment
of the
Requirements for the Degree

DOCTOR OF EDUCATION

ATHENS, GEORGIA
1974

ACKNOWLEDGEMENTS

My sincere thanks go to my professors who so ably served on my committee: Dr. W. Owen Scott; Dr. James Kenyon; Dr. Albert J. Kingston; Dr. Mary Hepburn; and Dr. Everett Keach. Praise and special thanks are due a man who for two years has shown me many of the finer things in learning. I refer to my Major Professor Marion Jennings Rice, gentleman and scholar.

Special appreciation is extended to:

Ross, Rob, Jim and those other graduate students who assisted in the development of the materials, statistical analysis, proofreading of the manuscript, and the stimulating environment in which I found myself.

Evelyn for the outstanding job of typing.

Acknowledgement is made to the following personnel for helping make this study possible:

Savannah-Chatham County School District, Field Testing

Superintendent: Mrs. Saxon P. Bargeron
Social Studies Supervisor: Mr. Bernard A. Hirshberg
Beach Junior High
Principal: Mr. R.S. DeLoach
Teachers: Mr. L. Passink
Mrs. P. Gunter
Mercer Junior High
Principal: Mr. R.L. Rush
Teacher: Miss S. Chamberlin
Shuman Junior High
Principal: Mr. R.A. Bryant
Teacher: Mr. M. Lariscy
Myers Junior High
Principal: Mr. G.P. DeMott
Teacher: Mrs. A. DeVeaux

FOREWORD

This study was undertaken as part of the continuing research and development of the Geography Curriculum Project, University of Georgia.

The content focus of the Geography Curriculum Project is the preparation of supplementary units for the elementary grades, emphasizing the organizing concepts of the discipline of geography. The research focus is the testing of some psychological construct of learning, such as the nature of concepts, Ausubel's reception learning model, Bloom's mastery learning, or Bruner's discovery hypothesis, under normal conditions of school instruction.

The Geography Curriculum Project thus serves as a small research and development center. It develops new materials and measurement instruments, field tests and evaluates materials, and facilitates the training of doctoral students in geographic education.

The Geography Curriculum Project was initiated as a result of a study of geographic content in elementary social science texts, manuals, and study guides. The evidence indicated that elementary geography is primarily presented as a discrete body of facts, with little attention to the organizing concepts of geography which help to analyze, interpret, and integrate physical and cultural phenomena. The development of systematic geography units helps to clarify the teaching of geographic knowledge and concepts. The research emphasis answers questions relating to the structuring of materials and their use in teaching geography.

DEDICATED TO MY WIFE

LORNA

AND OUR DAUGHTER

JENNIFER

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENTS iii

FOREWORD v

LIST OF TABLES xi

CHAPTER

I BACKGROUND TO THE STUDY 1

General Statement of the Problem 6
Definition of Terms 7

II REVIEW OF THE LITERATURE 11

Antecedents of Mastery Learning 12
Research Related to Relevant Mastery Learning Variables 16
Comparisons of Learning by Mastery and Non-Mastery Procedures 17
Summary 28
Mastery Learning and Aptitude 28
Summary 31
Mastery Learning and Retention 32
Summary 37
Mastery Learning and Times-to-Testing 37
Summary 40
Mastery Learning and the Social Sciences 41
Summary 44
Conclusion 44

III DEVELOPMENT OF MATERIALS USED IN THE STUDY 46

Construction of the Curriculum Materials 46
Text Content 46
Chapter Format 48
Workbook Content 50
Workbook Format 50

CHAPTER Page

Treatment Procedures. 51
  Procedures Common to Both Treatments. 53
  Additional Procedures for Mastery Learning. 54
  Readability 55
Construction and Characteristics of the
  Review and Summative Achievement Tests. 57
  Construction of the Review Tests. 57
  Content Validity. 58
  Realiability. 59
  Construction of the Summative Test. 59
  Content Validity. 60
  Reliability 61
Alterations in the Treatment Procedures 63
Summary 66

IV METHODOLOGIES AND PROCEDURES. 68

Experimental Design 68
  Rationale for the Design. 68
  Rationale for the Concomitant Variable. 71
  Unit of Statistical Analysis. 74
Experimental Study. 74
  Sample Selection. 74
  Random Assignment of Individuals
    to Treatment Groups 74
  Distribution of Students by
    Treatment and Aptitude. 76
  Reading Scores, Grade Equivalents, and
    National Percentile Rank. 77
  Orientation of Teachers 79
  Duration of the Study 80
Pattern of Logic Used in the Study. 80
  Research Hypotheses 80
  Pattern of Logic for Testing the Research
    Hypothesis Statement. 81
  Discussion of Pattern of Logic. 82
Contextual Variables. 84
  Community and School District 84
  Characteristics of the Schools in
    the Study 84
  Characteristics of the Teachers in
    the Study 86
  Summary of Contextual Variables 87
Statistical Procedures. 88
  Statement of the Statistical Hypotheses 89
  Significance Level. 93

CHAPTER Page

Assumptions Underlying the Multivariate Analysis of Variance (MANOVA). 94
Assumptions Underlying the Analysis of Variance (ANOVA) 94
Limitations. 95
Summary. 98

V RESULTS AND DISCUSSION OF THE FINDINGS 100

Presentation of the Findings 100
Analysis of the Data by the Multivariate Analysis of Variance (MANOVA). 100
Findings of Statistical Hypotheses 100
Discussion of the Findings 118

VI SUMMARY, DISCUSSIONS, AND RECOMMENDATIONS 125

Summary. 125
Research Hypotheses. 125
Procedures 126
Findings 128
Findings of the Treatment by Aptitude Interaction. 128
Findings Between Treatment Groups. 128
Findings of the Aptitude Levels Across Treatment: Simple Effects 128
Discussion of Educational Implications 129
Recommendations for Further Research 131
Summary of Recommendations 133

REFERENCES. 135

APPENDICES 142

Appendix A - Student Text: Functions of Cities used by both Mastery and Non-Mastery Treatment Groups. . . . 142
Appendix B - Student Workbook for the Non-Mastery Treatment Group 228
Appendix C - Student Workbook for the Mastery Treatment Group 331
Appendix D - List of Major Facts and Concepts to be Learned 462
Appendix E - Table of Specifications for Achievement Tests 464
Appendix F - The 40 Item Multiple Choice and the 24 Item Recall Geography Achievement Test. 467

Page

Appendix G - Report from Teachers' Weekly Report Form. 479
Appendix H - School Correspondence Forms. 481

LIST OF TABLES

TABLE		Page
2.1	Summary of the Number of Mastery Learning Research by Content Areas and Level.	16
2.2	Studies Comparing the Results of Mastery Learning and Non-Mastery Learning Procedures	18
2.3	Studies Indicating the Effects of Mastery Learning Upon Aptitude Groups.	30
2.4	Studies Indicating the Effects of Mastery Learning Upon Retention	33
2.5	Studies Correlating Achievement and Time Spent in Learning	38
2.6	Studies in the Social Sciences Using a Mastery Learning Procedure	42
3.1	Cities by Geographic Location and Function	47
3.2	Reading Ease Scores and Grade Level for 17 Samples Selected from Nine Chapters of the Materials Functions of Cities	58
3.3	Test Analysis Data for the 40 Item Multiple Choice Instrument.	62
3.4	Test Analysis Data for the 24 Item Recall Instrument	62
3.5	Mean Scores for the 40 Item, 24 Item, and Total Achievement Measure by Treatment and Aptitude	64
3.6	Percentage Correct for the 40 Item and 24 Item Achievement Measure by Treatment and Aptitude	65
4.1	Experimental Layout (MANOVA)	69

TABLE Page

4.2 Number of Students by Reading Level by Class and Treatment Including those Students Omitted from Data Analysis. 75

4.3 Mean Reading Scores, Standard Deviations, and Grade Equivalents by Treatment and Aptitude Level 78

4.4 Mean Reading Scores, Standard Deviations, and Grade Equivelents for Aptitude Groups 79

4.5 Mean Reading Grade Equivalents and National Percentile Ranks for Aptitude Groups. 79

4.6 Teachers and the Number of Classes Taught. 88

5.1 Learning Cell Mean Scores by Treatment, Aptitude, and Class 101

5.2 Retention: Cell Mean Scores by Treatment, Aptitude, and Class 102

5.3 Times-to-Testing and Amount Completed: Cell Mean Scores by Treatment, Aptitude, and Class. 103

5.4 Multivariate Analysis of Variance Test of Significance - Learning, Retention, and Times-to-Testing 104

5.5 Analysis of Variance for Treatment, Aptitude, and Interaction - Learning. 104

5.6 Analysis of Variance for Treatment, Aptitude, and Interaction - Retention 105

5.7 Analysis of Variance for Treatment, Aptitude, and Interaction - Times-to-Testing. 105

5.8 Learning Mean Scores by Aptitude 108

5.9 Learning: Summary of Results of the Duncan Multiple Range Test at the .05 Level of Significance for Aptitude Effect 109

5.10 Retention: Mean Scores for Treatments 110

5.11 Retention: Mean Scores by Aptitude. 111

TABLE Page

5.12 Retention: Summary of Results of the Duncan Multiple Range Test at the .05 Level of Significance for Aptitude Effect. 111

5.13 Times-to-Testing in Minutes : Mean Minutes for Treatments 112

5.14 Learning: Cell Means and Differences for Treatments Across Each Level of Aptitude. 114

5.15 Retention: Cell Means and Differences for Treatments Across Each Level of Aptitude. 116

5.16 Times-to-Testing: Cell Means in Minutes and Differences for Treatments Across Each Level of Aptitude 117

5.17 Summary of Multivariate and Univariate Tests of Significance: Interaction and Main Effects 119

5.18 Summary of Tests of Significance for Simple Effects: Comparison of Aptitude Levels Across Treatments. . . 120

CHAPTER I

BACKGROUND TO THE STUDY

A continuing educational challenge is how to organize instruction in schools to facilitate a high level of learning for the majority of students. This problem is the core of pedagogy - how to help students learn more in a given time - and might be regarded as the departure point for the development of a science of instruction and learning.

In the decade of the 60s, this challenge of organizing instruction to facilitate learning assumed a new urgency with the re-discovery of the disadvantaged learner. Under the slogan of "compensatory education," a variety of programmatic attempts have been made to overcome the learning deficits of the slow learner, especially learning deficits which might be attributed to a disadvantaged environmental background.

Success in school subjects is now regarded not merely as a matter of school achievement but of personality and social adjustment as well. Low school performance is cumulative. Consequently, low performing students are seldom able to overcome learning deficits. Continual low performance reduces a student's desire for further learning (Sears, 1940) and develops undesirable attitudes toward learning (Khan, 1969). In turn, these traits lead to the development of poor self-concept (Torshen, 1969) and possible mental health problems (Stringer and Glidewell, 1967). Some critics, such as Block (1971) allege that as few as one-third of the students have successful and rewarding learning experiences under traditional assign-recite-test procedures adapted to the class mean.

Mastery learning has been proposed as a teaching-learning procedure that may substantially increase the proportion of students enjoying successful and rewarding school learning experiences. Mastery learning is a term coined by Bloom who contends that ".... all or almost all students can master what they are taught." In contrast to programmed instruction designed for individual self-instruction, feedback, and re-learning, Bloom's mastery learning envisions the use of procedures "whereby each student's instruction and learning can be managed *within the context of ordinary group-based classroom instruction*, as to promote his fullest development."

Bloom not only proposes mastery learning as an alternative which will give lower performing students the necessary additional time to learn, but he even alleges that mastery procedures will minimize differences in achievement resulting from differences in aptitude. He claims that as many as ninety-five per cent of the school population can learn most of the material to a stipulated criterion level provided they are given sufficient time and adequate correction and feedback. Mastery procedures will not be effective for five per cent of the population because of innate learning disabilities (Bloom, 1968).

The Bloom hypothesis that mastery learning procedures can overcome aptitude differences is contrary to the mass of psychological evidence which indicates that most treatments are insufficient to overcome differences in aptitude (DeCecco, 1968) and that methods of teaching share the common result of ineffectiveness (Wallen and Travers, 1963). In the research of the Georgia Anthropology and Geography Curriculum Projects, the evidence consistently shows that aptitude, as measured by reading test scores, is a more significant learning variable than methods of treatment (Steinbrink, 1970; Frech, 1973; Dumbleton, 1973).

Furthermore, Bloom mastery procedures are class-paced rather than individual-paced mastery. In the Bloom procedure, the progress of the higher aptitude student is retarded by the withholding of additional learning tasks. Instead, he serves as a tutor or teacher aide to assist the lower performing and slower student. In contrast, in individual-paced instruction, whether of the earlier Winnetka type (Washburne, 1922) or the more recent IPI-type (Glaser, 1968), the higher aptitude student has consistently achieved at a higher performance level and completed more units of study.

In a class-paced mastery procedure, as proposed by Bloom, low achieving students attain the criterion level attained by high achieving students. But the increase in achievement by low aptitude students is attained at the cost of two trade-offs which may not be educationally desirable. One is the slow down in the achievement pace of the high aptitude student. This use of high aptitude talent to assist low achievers might, in the long run, constitute a waste of educational talent. The short-term run of most mastery studies thus far, however, neither provide the evidence for the abuse of high aptitude student talent nor the long-term efficacy of mastery procedures for low aptitude students.

The second trade off is in the amount of time required to attain the criterion level established for "mastery." The provision of extra learning time for the low aptitude student may provide a substantial learning difference.

One of the alleged advantages of mastery is that while the procedure may be initially slower, the thorough learning of content and procedures facilitates subsequent learning. This claim may hold some merit for hierarchically organized subjects, such as mathematics or foreign languages, but may not be true for subjects, such as the social sciences, in which the complexity of the subject matter appears to be primarily a function of factual, conceptual, and syntactical complexity rather than the sequencing of learning hierarchies.

The social studies contain learning clusters based on the concepts and facts being presented, but their sequencing, however logical, appears to be arbitrary. For example, in both the Anthropology and Geography Curriculum Projects at the University of Georgia several topical alternatives were considered in the sequencing of the content. In mathematics, foreign languages, accounting, and shorthand, in contrast, there are generally agreed on progressions of presentation moving from the simple to the more complex. Mastery procedures may facilitate subsequent learning in elementary arithemetic but mastery procedures may not transfer to elementary history, because new factual and conceptual material is largely discrete.

Thus in the social studies it might be possible to attain mastery over a portion of the material to be covered, but this intensive coverage is attained at the expense of a more extensive treatment. Time to teach and learn in a school setting is limited. Consequently, it is not educationally desirable to ignore the amount of time required to achieve a given task. In the Carroll model of school learning (Carroll, 1963), aptitude is a function of the time taken to learn. Consequently, any investigation of mastery learning must take into account the time students take to achieve mastery. Time is thus not only a contextual variable, but it may also be regarded as an important treatment variable.

Research in mastery learning to this date has not systematically examined the various variables implicit in any learning system. Rice (1973) identified seven independent variables and four dependent variables which require systematic examination to establish a body of evidence to substantiate the allegations of mastery learning. Generally, mastery learning has been presented as a panacea (Block, 1971) with an overgeneralization and statement of claims. In a critical analysis of the state of the art and quality of research, Mitchell (1974, in draft) concluded that much mastery learning research is based on crude comparisons of a mastery group with a non-mastery group, often with ex post facto comparisons. Thus, while mastery learning procedures have generally been reported as superior to non-mastery procedures (Kim, 1969, 1970; Block, 1970; Lee, 1971), it is extremely difficult to assess the results of such research. The reader is left with the feeling that many comparisons of mastery with non-mastery procedures are merely comparisons of superior with inferior instruction, or may result from the halo effect of experimental treatment.

In selecting a focal point for this study in mastery learning, it was decided to design a study which would give importance to the aptitude variable in mastery learning. This question appeared to be crucial, for, as the review of the literature in Chapter II shows, there appears to be a tendency to make claims for mastery learning which are not substantiated by the evidence.

General Statement of the Problem

The central question this study addresses itself to is this: If a mastery procedure is used in teaching a geography unit at the grade seven level, will the average achievement of students at three levels of aptitude be significantly different?

Three aptitude levels were arranged using the word meaning section of the Iowa Tests of Basic Skills: Forms 5 and 6 (Lindquist and Hieronymus, 1971) as the concomitant variable. A high, middle, and low group were formed. Since achievement may be measured in terms of learning, as assessed by immediate posttest, and by retention, as measured by a delayed posttest, it was decided to measure both learning and retention to see if mastery procedures might demonstrate superiority with a time interval in testing. The treatment consisted of a self-instructional geography text and workbook Functions of Cities, Publication No. 74-1, Geography Curriculum Project, University of Georgia.

In all teaching, the classroom unit of instruction appears to be crucial in educational research. Since educational researchers typically must use intact classes rather than randomize assignment of students to treatment, the research design must take into account the classroom and teacher variable. In order to minimize teacher effect, it was decided to use self-instructional materials. But since students work differently under different teachers, it was considered necessary that the data analysis take into account the classroom variable.

A second aspect of the Bloom hypothesis implicity relates to the time variable. Given enough time and proper feedback, mastery procedures allegedly overcome aptitude differences. But if high aptitude students are able to continue to work at learning tasks, not limited to tasks which are paced to the slower learner, would not higher aptitude students not only cover more, but achieve at a higher level?

Definition of Terms

For the purposes of this study, the following terms were used:

Mastery Learning is used in accordance with general usage to describe a teach-test-reteach strategy. There are no set procedures for mastery learning. There are two major patterns--group-paced, sometimes called the Bloom model (1968), and individual paced, sometimes called the Keller model (1968). The operational characteristics, however, of any mastery treatment vary with the procedures stipulated by the investigator. In this study, the mastery procedures include diagnosis, correction, and restudy after the administration of two review tests. After completion of the second review procedure, the mastery students were permitted to continue to the next unit, even without attaining the criterion. Since the operational procedures are discussed at length in Chapter III, pp. 51-54, the specific procedures will not be developed at this point.

Non-mastery learning is a general term used to describe teaching-learning procedures which do not provide systematic feedback and opportunity for a student to restudy and learn the subject matter to a specified criterion. Any kind of instructional procedure, group or individual, class paced or personalized, structured or unstructured, open or closed, may be used as a non-mastery procedure.

In this study, non-mastery procedures include the use of a structured text with accompanying workbook, review test, and class discussion. These procedures, as described in Chapter III, are part of the self-instruction also administered to students in the mastery group. In order for a comparison of mastery and non-mastery procedures to be carried out each procedure must be carefully designed and adhered to. In addition, the content should be identical. The only differences in the organization of the content should be those differences which are essential in making the treatments distinct. The critical difference in the mastery and non-mastery treatments, as stipulated in this study, is the requirement that mastery students restudy material and attain a specified criterion, 85 per cent, before proceeding to the next unit. The non-mastery treatment, in contrast, does not provide additional time for restudy and learning.

Aptitude, in this study, was used to describe a level to which a student was assigned as measured indirectly by the word meaning section of the Iowa Tests of Basic Skills: Forms 5 and 6 (Lindquist and Hieronymus, 1971). It refers to a student's capacity or talent to learn or understand. Correlation of student performance and an indirect measure such as an aptitude word-meaning test have proven to be high (Thomas, 1967; Gaines, 1971; Dale, 1972; Pelletti, 1973), and as such, are good predictors of scholastic aptitude.

Learning is the knowledge and application of facts, concepts, and generalizations acquired as a result of study in one of the treatment groups as measured by a posttest directly related in content to the cognitive objectives of the materials, administered immediately upon conclusion of the treatment period. Knowledge, as used in this definition, is used in the general sense of knowing (Webster's Third New International Dictionary, 1971), and is not to be construed in the limited sense of knowledge implied by the Bloom taxonomy (Bloom, 1956).

Retention is the amount of knowledge retained as a result of studying in one of the treatment groups as measured by the same form of a posttest for learning administered as a delayed posttest.

Times-to-testing is the mean classroom elapsed time taken by students in each cell to complete or partially complete the treatment materials.

Criterion level is a score which mastery students must reach on a unit review test in order to proceed to the next unit. The eighty-five per cent level was used as the criterion level in this study. This criterion was selected because the studies of different criterion levels cited in Chapter III indicate that the 85 per cent level is sufficiently high to encourage a greater quality of learning, but not too high to be discouraging, especially to the lower aptitude student.

Review test is a test administered to each student at the completion of each chapter. Review tests were used as an indication of the quality of learning to students and as a reference for reviewing poor quality learning. Both mastery and non-mastery students completed the first review test but at its conclusion non-mastery students proceeded to the next chapter of work, while mastery students, who did not reach criterion, restudied the text and workbook exercises. When they finished restudying, they took a second review test. This test contained the original items, however the items were reordered. Chapter III, pp. 58-60 contains a more complete explanation of the review tests. The term 'review test' has been used in this study in lieu of the Bloom, Hasting and Madaus (1971) term of 'formative' evaluation. However, their meanings are not synonymous.

This discussion of terminology is pertinent to the review of the literature, the subject of the next chapter, and to the methods and procedures of writing the treatment materials and preparing the measuring instruments, presented in Chapter III.

CHAPTER II

REVIEW OF THE LITERATURE

The present study was designed to compare the average achievement levels of mastery and non-mastery procedures of high, middle, and low aptitude students, using measures of learning, retention, and times-to-testing. Students bring a wide range of aptitudes to each learning experience. It is the hope of teachers that students learn and retain learning to a high degree. A teaching-learning procedure that facilitates the learning expectations of teachers for students of varying aptitudes would offer a valuable contribution to education. However, if such a procedure were to require more learning time the economics of class learning interacting with the many school subjects might be disadvantageous.

Three independent variables were used in this study. They were 1) treatment (mastery and non-mastery); 2) aptitude (high, middle, and low); and 3) class (10 classes for treatments). Three dependent variables were used. They were: 1) learning (Geography Achievement Test, posttest); retention (delayed posttest); and 3) times-to-testing (elapsed classroom time).

The discussion of the literature will focus on the independent and dependent variables to be used in this study. Therefore, the following organization was used: 1) antecedents of mastery learning; 2) comparisons of learning by mastery and non-mastery procedures; 3) aptitude; 4) retention; 5) times-to-testing; and 6) mastery learning and the social sciences.

Antecedents of Mastery Learning

Very few ideas in education today are without a firm base in earlier pedogogy. Benjamin Bloom's (1968) mastery learning strategy is no exception. Prior to Bloom's publication several notable attempts were made in the United States to develop systematic teaching-learning strategies. Among the systems devised were those of Washburne, Morrison, and Skinner. However, it was Carroll's (1963) Model of School Learning that provided the theoretical background for the concept of "mastery." Washburne's (1922) work with the Winnetka School System in Chicago was one of the first of note. The Winnetka Plan aimed to individualize pupil instruction by building a curriculum in which time was varied and achievement was constant. This required that subject matter objectives be clearly stated, instructional materials be sequential, appropriate criterion levels be fixed, diagnostic-progress tests be constructed, and supplementary self-instructional materials be designed.

The results of experiments conducted at Winnetka indicate that pupils in the individualized program did not achieve any higher than pupils in conventional classrooms. However, the individualized program did appear to reduce the amount of time the pupils spent in learning (Washburne, Vogel and Gray, 1926).

Morrison (1926) developed a strategy similar to that of Washburne using students at the Laboratory School of the University of Chicago. He developed the strategy of, "Pre-test, teach, test the result, adapt procedures, teach and test again to the point of actual learning, (p. 79)." The Morrison model was based on the premise that learning was attainable given enough time and proper instruction. Morrison stressed that reteaching procedures should reflect careful decision making on the part of the teacher after he had reviewed the results of student tests. The Morrison approach specifically called for test results to act as the focusing agent for both student and teacher when further instruction was under consideration.

Washburne's and Morrison's strategies did not appear to be favorably received within the field of social studies. Boyington (1932) and Boten (1932) contributed the only reported research found in the field. They developed diagnostic tests for detecting weaknesses in the teaching and learning of social studies content. It would appear at this juncture that the strategies developed by Washburne and Morrison did not achieve favor due to the development of other strategies, such as problem solving.

The "teach, test, reteach" strategy did not resurface until the late 1950s and early 1960s. Skinner (1954) revived them through his development of programmed instruction. The principal idea of programmed instruction was that learning of any behavior, no matter how complex, rested upon the learning of a sequence of less complex component behaviors. Programmed instruction operationalized Skinner's stimulus - response learning theory and it appeared to facilitate learning for those students who required small learning steps, drill, and frequent reinforcement. However, it did not facilitate learning for all or almost all students. Carroll's (1963) 'Model of School

Learning' attempted to fill this gap.

Essentially, Carroll's model was a conceptual paradigm that outlined factors influencing and interacting to produce student success in school learning. In its simplest form, his model proposed that if each student was allowed the time he needed to learn some stipulated criterion level and he spent the required learning time, then he could expect to attain that level. If the student was not allowed sufficient time, then the degree to which he could expect to learn was a function of the ratio of time actually spent in learning to time needed:

$$\text{Degree of learning} = \frac{(\text{time actually spent})}{(\text{time needed})}$$

Carroll's model conceived of school learning as consisting of a series of distinct learning tasks. In each task, the student proceeds ".... from ignorance of some specific fact or concept to knowledge or understanding of it or .... from incapability of performing some act to capability of performing it (Carroll, 1963, p. 723)." The model proposed that under typical school learning conditions, the time spent and the time needed were functions of certain characteristics of the individual and his instruction. The time spent was determined by the amount of time the student was willing to spend actively engaged in learning and the total learning time he was allowed. The learning time each student required was determined by his aptitude for the task, the quality of instruction, and the student's ability to understand instruction. These are the factors that specify the sources of variation that should be included in the model and which have been used as specified by the model. The Carroll model is a figurative model not a mathematic model, and as such, the components are not additive. The full Carroll model can now be summarized as:

$$\text{Degree of learning} = f\left[\frac{\text{1. Time allowed} \quad \text{2. Perseverance}}{\begin{array}{l}\text{3. Aptitude} \quad \text{4. Quality of instruction}\\ \text{5. Ability to understand instruction}\end{array}}\right]$$

Bloom (1968) transformed Carroll's conceptual model into a working strategy for mastery learning. The mastery learning strategy proposed by Bloom was designed for classrooms where the time allowed for learning was relatively fixed. Mastery was defined in terms of a specific set of major objectives the student was expected to exhibit by the end of a unit of classroom study. The content was then broken into a number of smaller learning units and the unit objectives were defined where criterion to mastery was essential for mastery of the major objectives. The instructor taught each unit using typical, group-based methods but supplemented this instruction with feedback-correction procedures to ensure that each student's unit instruction was of optimal quality. The feedback devices were brief review evaluations administered at unit completion. Each evaluation covered all objectives of a particular unit. Student achievement on the unit objectives indicated the level of each student's learning. Supplementary instructional correctives were then applied to help students overcome their unit learning problems before continuing with the group instruction.

Since 1968, when Bloom published the mastery learning paradigm, a number of compendiums have been compiled surveying the efficacy of mastery learning, both nationally and internationally (Block 1971 & 1973, Mitchell, in draft). Mastery learning has been implemented at many levels of education, but research predominates at the college level. Successful strategies have also been incorporated into subjects ranging from mathematics to psychology to physics (see Table 2.1).

Research Related to Relevant Mastery Learning Variables

Block (1973) indicates that there has been and continues to be a growing body of research that supports the use of mastery learning procedures across a broad spectrum of disciplines and levels. Figure 2.1 presents a selected summary of mastery learning research by content area and level that will be covered in this review.

Table 2.1

Summary of the Number of Mastery Learning Researchers by Content Areas and Level

Level	Math	Science	Psych.	Social Studies	Language	Other	Total
College	3	3*	3*	2*			11
High School						1	1
Junior High School	3**			1***	1**		5
Elementary	3			1***	1		5
Total	9	3	3	4	2	1	22

*The study of Moore, Mahan, and Ritts (1968) was conducted in three content areas.
**The study of Kim (1968) was conducted in two content areas.
***The study of Gaines (1971) was conducted at two levels.

The data of Figure 1 indicates that 50 per cent of the research has been conducted at the college level and that approximately 50 per cent has been focused upon mathematics. The remaining content areas and levels have not received as much attention. Of the studies reviewed, nine compare a mastery learning and a non-mastery learning procedure; three involve comparisons of mastery learning with aptitude; four include retention; two correlate achievement with time spent in learning; and two report results of mastery learning in social science disciplines. The review of these five sections follows.

Comparisons of Learning by Mastery and Non-Mastery Procedures

Nine research studies have compared mastery learning to non-mastery learning procedures. Table 2.2 provides a summary of the studies. Typically, these studies report results that use data obtained from a final cognitive summative achievement test.

Airasian (1967) applied a modified version of Carroll's model of school learning to a class (n=33) of graduate students in test theory. The objective was to facilitate mastery of the content for all students over a ten-week period. Ungraded formative evaluations were used to indicate strengths and weaknesses of student learning and instruction. Time inventories were tallied twice a week to determine the amount of time spent on study. Student achievement was measured by a summative test. His results indicate that, whereas during the previous year 30 per cent of the students received an A, 80 per cent of the sample achieved at or above the previous year's A grade score on a parallel exam and thus received A's.

Two other results were also of interest. First, the correlation

Table 2.2

Studies Comparing the Results of Mastery Learning and Non-Mastery Learning Procedures

Year	Study	Level and Subject Area	Number in Sample	Duration	Strategy	Results
1967	Airasian, P.W.	Graduate Test Theory	33	10 weeks	Non-graded Diagnostic Tests	80% achieved A compared to 30% previous year on parallel exam. Slight negative correlation between time and achievement. Less variability over time in achievement on formative evaluations.
1968	Mayo, S.T. Hunt, R.C. and Tremmel, F.	College Statistics	17	6 weeks	Feedback and formative tests	65% of experimental (mastery) group received A's while only 5% of the comparison group received A's.
1968	Moore, J.W. Mahan, J.M. and Ritts, C.A.	College Biology, Psychology, Philosophy	70	N.S.	Self-instructional tests and review	80% attained criterion in philosophy (mastery group) compared to 60% in previous year's class (non-mastery group). Experimental group mean was one-half standard deviation above control group mean on aptitude scores) for biology and psychology.

Table 2.2 (Cont'd)

Year	Study	Level and Subject Area	Number in Sample	Duration	Strategy	Results
1969	Collins, K. M.	College Algebra, Calculus	N_1=50 N_2=40	N.S.	Mastery including objectives, non-mastery	75% of students achieved criterion of A or B grade in mastery group against only 30% in non-mastery in algebra. 65% achieved to same criterion against 40% in calculus.
1969	Green, Ben A.	College Introductory physics	150	N.S.	Self-paced instructional units, formative evaluation, tutors, programmed review materials	Cognitive and affective outcomes of mastery group as positive group learning under traditional lecture-discussion-demonstration. Student tutors added human dimension to learning that the use of technological aids lacked.
1969	Kim, H.	7th grade math	273	N.S.	Objectives, formative evaluations, review questions, student tutors	74% attained mastery criterion whereas 40% attained criterion in non-mastery group. Students with 93 IQ, 50% attained mastery criterion, whereas 8% attained criterion in non-mastery group. High IQ group 95% attained mastery criterion, 64% in non-mastery group.

Table 2.2 (Cont'd)

Year	Study	Level and Subject Area	Number in Sample	Duration	Strategy	Results
1970	Biehler, R. F.	College Introductory Psychology	N.S.	N.S.	Course Objectives, unit tests with parallel form, term papers	Anecdotal results reported; strategy improved cognitive and affective outcomes.
1970	Gentile, J. R.	College Introductory Psychology	52	one semester	self-paced, small units	Significant difference between mastery and non-mastery students at .001.
1970	Kim, H.	Middle School Math, English	5800	8 weeks	Used a mastery format, diagnosis, formative evaluation, feedback-correction, summative test	75% gained mastery criterion level compared to 28% under ordinary instructional conditions in English, 61% mastery compared to 39% non-mastery in math.

between total hours of weekly study and achievement was slightly negative. Airasian suggests that this may have been due to the effectiveness of the feedback system in leveling initial differences in prior exposure to the course materials. It would appear that the diagnostic tests, by identifying important course aims and behaviours, facilitated positive student use of time. Second, there was less variability over time in achievement on the formative evaluation instruments. In spite of the varying backgrounds possessed by the students, this strategy appeared to be effective in bringing most of the students to a high degree of achievement by the end of the course. However, Airasian does not indicate how many students were repeaters from the previous year or whether repeating students may have provided an inflated result.

Mayo, Hunt, and Tremmel (1968) conducted a six-week university summer session in introductory statistics that emphasized the use of homework and weekly formative tests accompanied by individual and small group assistance. Student grades were assigned by student performance in class rather than by relative academic standing within the class. Both the mid-term and summative examination were used to produce a grade.

Seventeen students were assigned to either a mastery learning or a comparison group. The results indicate that 65 per cent of the mastery learning group received an A whereas only 5 per cent of the non-mastery group reached that standard. It was found that the feedback procedures (formative evaluations) and the tutoring facilitated student achievement in the mastery learning group.

In a study by Moore, Mahan, and Ritts (1968), students were presented self-instructional materials in biology, psychology, and philosophy. Students were tested at the conclusion of each unit (formative evaluations) and, if mastery was not achieved, they were redirected through additional instructional materials and alternative test forms until mastery was exhibited. The students were required to reach a predetermined achievement level that was equivalent to an A or B on the traditional grading system.

Students learning biology and psychology were divided into experimental and control groups (N=35 in each group). The results of the summative test indicated that the experimental group achieved approximately one-half standard deviation above the control group. For students in philosophy, the grades of the experimental group were compared to a control group from the previous year. Approximately 80 per cent of the experimental group received an A or B compared to 60 per cent of the control group. These results should be treated carefully due to the reporting technique used. Weak research design and statistical analyses should not be used to make even moderate inferences about a treatment. This dictum appears to have been violated in this study.

An investigation of the effectiveness of Bloom's mastery learning strategy for teaching a freshmen college mathematics course was conducted by Collins (1969). Two algebra courses for liberal arts majors were used. Students were assigned to a mastery learning and a non-mastery group.

The mastery learning group was given a list of course objectives to be covered in each unit, each class session, and each assignment. During each class session, up to ten minutes was allowed to solve a problem based upon the objectives from the previous session and assignment. The problem was then discussed and questions answered. Non-mastery learners received neither the objectives nor the daily problems. Both groups used the same textbook, covered the same material in class, and took the same summative test.

In the algebra classes, 75 per cent of the mastery compared to 30 per cent of the non-mastery students achieved the criterion of an A or B grade. In the calculus classes, 65 per cent of the mastery compared to 40 per cent of the non-mastery students achieved the criterion grades. In the mastery groups for both algebra and calculus, D and F grades were practically eliminated. The smaller differences in the percentages of students who attained the criterion under mastery and non-mastery learning conditions for the calculus courses may be attributed to three factors: (a) the greater importance of the courses to all engineering and science students; (b) the higher and more homogeneous mathematical ability of the calculus students; and (c) the clearer relationship between the problems discussed in class and the unit test problems.

Green (1969) used a mastery learning approach with 150 undergraduate students in teaching an introductory physics course. He used self-paced instructional units with formative evaluations, tutors, and programmed review materials. The purposes of the study were to determine if this particular mastery learning approach facilitated student achievement and whether student enjoyment was affected.

The results indicated that achievement, as well as enjoyment of the course, was as great on the final exam as students who learned under the traditional lecture discussion demonstration approach. Green suggests that the use of student tutors rather than the use of technological aids added a personal-social dimension to student learning. It should be noted that no statistically significant results are reported.

Kim's (1969) experiment examined the effectiveness of Bloom's strategies for mastery learning in Seoul, Korea where classes are predominantly very large (usually one teacher to 70 students).

The research sample consisted of 272 seventh graders. Half were assigned to the mastery learning (experimental) group and half to the non-mastery learning (control) group. These groups were comparable in terms of I.Q. and prior mathematics achievement. Both groups were taught a unit on simple geometric figures for eight sessions by their own teachers.

The results indicate that 74 per cent of the experimental compared to only 40 per cent of the control students attained the mastery criterion of at least 80 per cent correct answers on the summative achievement test. The data also reveal an interesting relationship between I.Q. and achievement under mastery and non-mastery learning conditions. Of those with below-average I.Q. (93), 50 per cent of the experimental students compared to only 8 per cent of the control students achieved the mastery criterion. Of those with above-average I.Q., 95 per cent of the experimental students reached the criterion compared to only 64 per cent of the control students. Thus, almost as many mastery students with below-average I.Q. reached the criterion as control students with above-average I.Q. Mastery learning appeared most effective for students with below-average I.Q.

A mastery learning strategy for teaching introductory undergraduate educational psychology was reported by Biehler (1970). The purpose of the strategy was to reduce examination pressure and competition among students through frequent test reinforcement. Students were allowed to select a traditional or mastery learning treatment group.

The mastery learning option contained a list of course objectives which was produced and circulated to each student. The list served as a basis for the construction of three normatively graded unit tests. Mastery performance was guaged at the cutoff for the ordinary A or B grade score levels. Students who failed to reach mastery performance reviewed the material and took an alternative test form. Three short papers and a term paper were also required. Final grades were assessed on the basis of mastery/non-mastery on the unit test and the writing of acceptable papers.

No statistical analysis of the data was attempted but through survey reporting Biehler suggested that students who performed poorly on the initial examination did not give up due to the procedure allowing alternative relearning procedures. These results are suspect, however, because of the subjective reporting technique.

Gentile (1970) describes a mastery approach to the teaching of a college course in introductory educational psychology. The purposes were to guarantee that all students mastered the main concepts; to demonstrate how instruction emphasizing cooperation rather than competition could be organized in the classroom; and to maximize interactions between students, student proctors, and the teacher. Student learning was self-paced over small instructional units. Study questions were provided to each student; student proctors (students who had already mastered the material) provided reinforcement and preparation for the unit test. If mastery was not achieved, the student was asked to review the material and then return for retesting. Proctors and the instructor were available at all times to help students review material. Each student who mastered all the units received an A.

The results of the mastery treatment were compared to a similar course more conventionally taught through large group, required lectures, and smaller discussion group sessions. The mastery approach produced significantly better understanding ($p<.001$) of comparable material taught in both courses. On identical forms of the course evaluation sheet, 74 per cent of the mastery students compared to 21 per cent of the control students indicated they enjoyed taking the course.

The achievement gains in this study must be called into question. Gentile indicated that "comparable" material was used with the control group. The failure to use the same treatment materials introduces a confounding variable that is difficult to control for, and hence, must influence generalizations based on the results.

In a later experiment, Kim (1970) reported the results of a large-scale expansion of his earlier experiment in mastery learning. Nine middle schools (approximately 5,800 seventh graders) in Seoul, Korea, participated. The experiment covered eight weeks of learning in mathematics and English.

Instructional strategies adopted in this project were much the same as those used in the first study (Kim, 1969), except that a diagnostic test to detect learning deficiencies and the necessary compensatory programmed units were administered prior to the regular instructional sessions.

The results indicate that the percentage of experimental students attaining mastery (80 per cent correct scores on the final summative examinations) varied widely across the sample schools. On the average, however, 72 per cent of the students reached the mastery criterion by learning English under experimental conditions compared to only 28 per cent learning under standard instructional conditions. In mathematics, an average of 61 per cent of the mastery compared to 39 per cent of the non-mastery students attained the summative achievement test criterion. Two schools did not follow the prescribed procedures. If the results for these schools are ignored, then 75 per cent of the mastery students attained the criterion level in English and 67 per cent in mathematics.

Fluctuations from school to school in the percentage of experimental students attaining the mastery criterion appear to have been caused by a) variation in school learning climate, b) variations in the school and teacher cooperation, and c) inefficient utilization and administration of the instructional materials. The school and the teacher are often variables that are overlooked in research. This experiment points up the importance of gaining full support and cooperation from the school and its teaching and administrative staff.

Naturally the findings of Kim must be interpreted carefully by instructional researchers. Korea is not the United States. However, we cannot afford to dismiss his results. He has used large numbers of students in order to support significant cognitive gains; hence, statistical differences may have been due to the large number of subjects used in the study and not necessarily to the effects of the treatment.

Summary

The nine studies which compared mastery with non-mastery support the idea that mastery procedures facilitates learning significantly more than non-mastery or control procedures. This finding is predictable in that a new or novel classroom learning mode will often find statistical significance in a classroom when compared to a traditional mode.

All studies reported here were conducted in intact classrooms. Individual differences were minimized through a variety of class-paced and individualized mastery learning techniques. However, the use of the classroom as the unit of statistical analysis does not provide strong support for studying characteristics such as the individual, the teacher, or the classroom learning environment. Consequently, other independent variables need to be isolated and evaluated.

The traditionally arranged classroom contains students who possess varying aptitudes to learn. An important question is whether mastery learning facilitates student achievement equally for all students or whether students with certain aptitude can benefit more from exposure to a mastery learning procedure.

Mastery Learning and Aptitude

Two studies have been located that specifically research the effects of mastery learning on students of varying aptitude (Table 2.3). The criteria for determining aptitude range from using I.Q. scores to selecting advantaged and disadvantaged socio-economic populations.

Carroll and Spearritt (1967) used 208 grade six students to observe relationships of intelligence and quality of instruction to achievement. Treatment was provided by self-instructional booklets, containing rules about verbs of an artificial language. The booklets differed in their presentation of the rules and in the amount of explanation of mistakes. Form A, the high quality of instruction form, presented each rule, tested it before presentation of subsequent rules, and referred the student to pages on which his mistakes were explained. Form B, the low quality of instruction form, presented a large quantity of disorganized information. The explanation of mistakes was also inadequate. Measures of learning rate, achievement, interest, and perseverance were administered.

This study determined that poor quality instruction depressed the performance of students at all the intelligence levels. However, there was an interaction between intelligence and the quality of instruction with respect to the student's willingness to persevere on a difficult post-experimental task. Students in the high and low intelligence groups who used the structured materials spent more time on the task than students in the middle intelligence group. Since, in this study, the average intelligence students applied themselves more to the post-experimental task their perseverance increased. However, the researchers speculated that poor quality of instruction decreased perseverance for the high and low intelligence students. A further finding was that

Table 2.3

Studies Indicating the Effects of Mastery Learning Upon Aptitude Groups

Year	Study	Level and Subject Area	Number in Sample	Duration	Strategy	Results
1967	Carroll, J.B. and Spearritt, D.	6th grade Artificial language	208 3 groups H M IQ L	N.S.	Two forms: A - highly structured B - disorganized Self-Instructional	Poor quality instruction depressed performance of children at all IQ levels; reduced perseverance of high IQ students; learning highly inefficient when insufficient opportunity to learn.
1970	Kersh, M.E.	5th grade Arithmetic	12 classes 6 advantaged 6 disadvantaged	1 year	Teacher instruction with feedback and review. Block model used.	Advantaged 1967 - 75% reached criterion 1966 - 19% reached criterion Disadvantaged 1967 - 20% reached criterion 1966 - 0% reached criterion increases in advantaged and disadvantaged groups under mastery learning procedure.

learning was inefficient when students had insufficient opportunity to learn, particularly where the instructional quality was poor and students were of low intelligence.

Kersh (1970) developed a mastery learning procedure based on the Carroll model and applied it to a unit in fifth-grade arithmetic. The unit was taught to six "advantaged" and six "disadvantaged" classes. This experiment has been reported in connection with the effects of a mastery learning procedure upon student retention (see Table 2.4).

The results of this study indicate that on the same achievement test and using the same mastery standards, there were significant increases in the proportion of experimental students (mastery class) attaining mastery compared to the proportion of students (control class) who attained mastery from the previous year. These increases ranged for one advantaged class from 19 per cent in the 1966 control class to 75 per cent mastery in the 1967 mastery learning class. The same teachers were used in both years, moreover, a disadvantaged class increased from 0 per cent attaining mastery in 1966 to 20 per cent attaining it in the 1967 mastery learning class. This may be an indication that the mastery learning procedure might be helpful in at least partially overcoming the cumulative deficit in learning apparently manifested in socio-economically disadvantaged students.

Summary

The question is, does mastery overcome the learning difficulties of students with varying aptitudes?

Aptitude is a personal quality of the learner. Therefore, it is questionable whether intact classes (Kersh, 1970) should be used as the unit of statistical analysis. Class mastery, by the Bloom hypothesis, retards the high aptitude student in an effort to advance the low aptitude student. Consequently, class mastery may well be contrary to the principle of individual instruction.

This review of mastery learning and aptitude variation determines that the original question has not been firmly answered by the evidence provided. It would appear, therefore, that a study that manipulates aptitude levels with other learning variables is required at this time.

Mastery Learning and Retention

Brownell (1948) refers to retention as the maintenance of skills or knowledge with no practice after the completion of the learning. Four studies within the mastery learning paradigm have been located that focus on retention as a variable (see Table 2.4). These studies tend to demonstrate the superiority of the mastery learning approach but they are not definitive.

Block (1970) established two tasks for his study. First, a rationale for setting objective, criterion-referenced performance standards for sequential learning tasks was proposed, applied and validated; second, cognitive and affective consequences of requiring students to maintain particular mastery levels throughout the learning of a sequential task were examined.

Three sequential units of elementary matrix algebra were taught to ninety-one eighth graders over a school week. However, due to student recalcitrance 17 per cent of the sample were dropped during the study. Many of these students were of low aptitude. Students were randomly assigned to either a control treatment or one of four mastery

Table 2.4

Studies Indicating the Effects of Mastery
Learning Upon Retention

Year	Study	Level and Subject Area	Number in Sample	Duration	Strategy	Results
1970	Block, J.H.	8th grade Matrix Algebra	91	5 days (2 weeks: retention)	Feedback-correction, formative evaluations, summative test. Used Bloom group-based model.	Linear relationship between amount learned and retention as measured by parallel test form. Only 85 and 95 per cent mastery group retained algebra to significantly greater extent than non-mastery group.
1970	Kersh, M.E.	5th grade Arithmetic	12 classes 6 advantaged 6 disadvantaged	1 year (summer holiday break	Teacher instruction with feedback and review. Used Bloom group-based model.	Positive significant effect in only one class. 63% of students in one advantaged class achieved criterion. Maintained 63% on retention test administration after summer break.

Table 2.4 (Cont'd)

Year	Study	Level and Subject Area	Number in Sample	Duration	Strategy	Results
1970	Romberg, A., Shepler, J., and King, I.	6th grade Mathematical proof, probability, statistics	N.S.	Retention: 2 weeks later mathematical proof 4 weeks later probability and statistics	Bloom type strategy used	Correlation between achievement and retention +.75 and +.78 on each unit respectively. Mastery students retained significantly more material than non-mastery students.
1970	Wentling, T.L.	High School Automobile Mechanics	N.S.	N.S.	Bloom mastery procedure and non-mastery procedure	Mastery treatments yield significantly greater scores than non-mastery on retention measure.

treatments. The control group learned algebra at their own pace with no criterion level required, but the mastery treatment groups were required to exhibit criterion performance on one unit before proceeding to the next. Each of the mastery groups were required to learn a different percentage of the material - either 65, 75, 85, or 95 per cent.

The findings indicate that there was a linear relationship between the percentage of material learned per unit and student retention as measured on a parallel-form of the summative achievement test administered two weeks after the close of instruction, that is, the higher the level to which each unit was learned, the greater the retention. However, only those students learning to the 85 and 95 per cent criterion retained the algebra to a significantly greater extent than the non-mastery treatment group. These results must be viewed tentatively as the number of students in the treatment groups was small, thus limiting the scope of the study. Moreover, it is difficult to ascertain when a retention measure should be administered to measure retention effectively.

In Kersh's (1970) study, six classes of fifth-grade students from socio-economically advantaged backgrounds and six classes from socio-economically disadvantaged backgrounds were taught arithmetic by their regular teacher over a full school year. The mid-year and end of year performance of these students was then compared with the mid-year and end of year performance of equivalent classes from the previous year. Further, students in the experimental classes were retested with a parallel form of the final exam at the beginning of the sixth grade.

The results of this study are equivocal. Because of teacher inability to follow the experimental procedures, only one class produced significant gains in achievement and retention. Sixty-three per cent of the mastery learning students still achieved to the 80 per cent criterion on the retention test administered at the beginning of grade six.

Kersh's study may be suspect due to the researcher's inability to control the teacher factor. This point reinforces a similar point made concerning Kim's (1970) study, that is, the necessity to control contextual variables especially classroom variables.

Romberg, Shepler, and King (1970) have reported results similar to both Block and Kersh. They had previously taught sixth grade students one unit of mathematical proof and another unit of probability and statistics. Students were expected to learn to a 90 per cent criterion level. Two weeks after the end of instruction, the students were given a delayed posttest using the identical form of the unit final examination.

Romberg and his colleagues found that the correlation between achievement and retention was .75 and .78 for the proof and probability units, respectively. The individual retention ratios, i.e., amount retained/amount learned, were approximately .95 for both units. It was also found that the mastery learning students exhibited significantly greater retention of the material learned than a matched group of non-mastery learning students.

In evaluating these results, it must be remembered that the same items appeared on the posttest and the retention measure. If students were given feedback about their posttest performance, then this feedback might have inflated the retention results.

In an unusual experiment, Wentling (In Press) taught a group of high school students a unit on automobile ignition systems as part of a course in automobile mechanics. Half of the students learned under a mastery learning strategy and half learned under a non-mastery learning strategy. The two types of instructional strategy were then crossed with two levels of intelligence and three feedback conditions. The data indicated that the mastery treatments yielded significantly greater scores than the non-mastery treatments on a retention measure administered after an undisclosed time at the end of instruction.

A flaw in this study occurred when Wentling used a feedback condition under both types of instruction which was probably an error. Accordingly, the retention data reported within the cells were confounded, which may have produced spurious retention results.

Summary

The mastery studies reported for the dependent variable, retention, have all been conducted in a Bloom class-paced situation. No retention studies have been found that have used the individual as the unit of analysis. Block (1973) concedes that more definitive research is required to determine whether individuals who learned to mastery retained more material than individuals who learned under non-mastery conditions.

Mastery Learning and Times-to-Testing

Few studies have examined the decremental and incremental effects of a mastery learning procedure upon time spent in learning (Table 2.5).

Table 2.5

Studies Correlating Achievement and Time Spent in Learning

Year	Study	Level and Subject Area	Number in Sample	Duration	Strategy	Results
1970	Merrill, M.D., Barton, K., and Wood, L.E.	College Imaginary Science	40	5 lesson span (3 weeks: retention)	Specific review of mislearned material, quizzes	Time in learning decreased as a result of specific review procedure. Even though experimental students received more material than the control group, they took less time to complete.
1970	Block, J.H.	8th grade Matrix Algebra	91	5 days	Feedback-correction, formative evaluations, summative test. Bloom model used.	Efficiency increased with students in the 95% mastery group by the end of the third unit. Learned more material in same time.

Merrill, Barton, and Wood (1970) pursued this line of research when they examined the effectiveness of a procedure to facilitate student learning of a hierarchical learning task. It was proposed that specific review at each stage where difficulties were encountered in student learning of a task should facilitate learning at subsequent stages. Forty college students were randomly assigned to two groups to learn an imaginary science through a five-lesson teaching machine course. In the experimental group, a specific review, step-by-step explanation was employed to facilitate learning of mislearned material. The control group did not receive specific review. In both groups, each lesson was followed by a quiz with no feedback of results. Immediately following the five lessons and quizzes each student was administered a criterion test.

The findings indicate that specific review following difficulties made experimental student learning increasingly efficient. The total time spent on original learning by the experimental group decreased successively across the five lessons. Further, the total time spent by the experimental group to complete the five lessons and accompanying quizzes, including the specific review material, was slightly less than the time spent by the control group. In other words, the experimental students studied more material than the control students but took less total time to learn it.

Support for the conclusion of Merrill, Barton, and Wood was provided by the results of Block's (1970) experiment (see Mastery Learning and Retention). In this study, the average total amount of learning time spent by each group was broken into the time spent in textbook learning and the time spent in correction/review. Attention was focused on the time spent by each group in original, textbook learning. The data revealed that as student learning progressed from unit to unit, students who maintained the 95 per cent level spent less time in original learning than students who maintained the other levels. This was especially apparent by unit three. Students in the 95 per cent group spent approximately the same average learning time as the control group. Hence, even if learning efficiency was measured in terms of per unit time rather than total learning time, the maintenance of none of the required levels made student learning more efficient than it might have been.

However, when learning efficiency was defined as the ratio of the average amount of original learning per unit to the average amount of learning time per unit, then the maintenance of the 95 per cent level made pupil learning more efficient by unit three. Students required to maintain the 95 per cent level learned approximately 40 per cent more material from textbook unit three than the control students, but they spent roughly the same amount of time in original learning as the control students.

It should be noted that in both of these studies elapsed time was recorded rather than the time spent in actual learning. Students rarely utilize the complete amount of time allowed for each subject.

Summary

Both reported studies used class-paced sequential abstract content over a short learning period of time of five days. Students particularly the slower learners, did not work-on-task and possibly experience the accompanying frustration that a longer period of time may have imposed. Therefore, there is a strong need for a study that provides individual students, of varying aptitude, with sufficient time to attempt a structured learning task that is longer than five days. This would also provide an opportunity to observe the effects of the procedures over time using an individual-paced procedure rather than a class-paced procedure.

Mastery Learning and the Social Sciences

Most mastery learning research has focused upon those disciplines that lend themselves to sequencing and hierarchical arrangement. Math and science have been well represented. However, there has been and continues to be a notable lack of research dealing with disciplines falling under the rubric of social science. The compartmentalizing of subject matter materials of the social science disciplines is not so readily possible as it is in mathematics or science (see Table 2.1). However, two studies that use social science disciplines in a mastery context were located (Table 2.6).

Gaines (1971) conducted a study with students from the fifth, sixth, seventh, and eighth grades using Georgia Anthropology Curriculum materials. The purpose of his study was to test presumed relationships of certain variables in John B. Carroll's model of school learning using two mastery learning strategies. Achievement, interaction between quality of instruction and ability to understand instruction, and the correlation of ability to understand instruction and degree of learning for both strategies were the specific variables under

Table 2.6

Studies in the Social Sciences Using a Mastery Learning Procedure

Year	Study	Level and Subject Area	Number in Sample	Duration	Strategy	Results
1971	Gaines, G.	5th, 6th, 7th and 8th grade Anthropology	28 classes	5 week unit	Used text, objectives, formative evaluations, feedback-correction, summative test. Used Carroll variables.	Carroll's interaction hypothesis between quality of instruction and ability to understand instruction not sustained. Important to control for confounding variables.
1973	Tierney, M.L.	College History	45	one quarter	Compared correction-feedback procedures as developed by Bloom (1968) and Keller (1966)	No significant difference between the use of correction-feedback procedures and traditional method. Significant differences found when alternative instructional modes used with Bloom strategy, also on application and affective criteria when pretest administered.

consideration.

While Gaines theoretically outlined two different treatment strategies - a mastery strategy of formative, multiple choice tests, and the non-mastery strategy of a workbook with completion items - the actual differences between the strategies may not have been sufficient to produce significant achievement differences.

Gaines also encountered inconsistencies with the administration of the treatment materials in schools. However, the eighth grade comparison favored mastery treatment using formative tests and was found significant.

Tierney's (1973) study involved two comparisons. First, was the comparison of feedback/correction components of two mastery learning strategies and a traditional lecture-discussion strategy to determine whether they produced significantly greater student achievement and attitude toward learning than a traditional mode. Second, was the comparison of an alternative instructional mode and the redirection of students into an original stimulus mode. His sample consisted of forty-five volunteer college students enrolled in an upper division European History class.

The study found no significant differences for either the achievement or affective criterions on the first comparison. However, significant differences were indicated between the two different mastery correction procedures on the application section of the achievement criterion. The alternative instruction mode produced students more able to apply the course material than those students who were redirected to original learning material.

Summary

Mastery learning research in the areas of social science has been minimal. The two studies reported suggest that application of a mastery strategy to the social science disciplines is in the formative stages. Consequently, there is a real need for a systema[illegible] appraisal of learning and contextual variables associated with mastery learning as applied to the social science disciplines.

Conclusion

Mastery learning has many roots in earlier pedagogy. These antecedents assisted Bloom to conceptualize what has now been termed 'mastery learning'. Since 1968, when Bloom coined the phrase, there has been little systematic study of variables associated with mastery learning.

Empirical studies, comparing a mastery to a non-mastery procedure, predictably, show support across a wide range of content areas. However, there has been no systematic attempt to determine whether slow learning students benefit from constant correction and feedback or whether, as Bloom claims, mastery can induce learning for nearly all students, particularly, when students are in an individual-paced situation with sufficient time to complete each learning unit.

A measure of effective learning is retention. The length of time between end-of-instruction and administration of the retention measure would seem to be important. In two of the reported studies two weeks intervened, one did not report, while the fourth tested after a summer break. A retention measure administered during the same academic year, and with a longer intervening time interval could lend stronger support to the mastery paradigm, particularly if slow learning students are involved.

The studies using time show that more efficient and economical use of school time might be expected. However, duration of the treatment may be a vital factor. Both studies ran for five days. Consequently, the question of whether a mastery procedure can save learning time or increase the amount covered or learned is not resolved.

Mastery learning research has been haphazard in its design and approach to independent and dependent variables. What is required is a systematic appraisal of the learning and contextual variables. This study is the first of a series that are planned to manipulate selected independent and dependent variables. The independent variables to be used in this study are aptitude, treatment, and class while the dependent variables are learning, retention, and times-to-testing. Social science materials have received scant attention from researchers. The use of geography material in this study within the mastery context, answers the call from Gaines (1971) and Tierney (1973) for application of a mastery procedure to disciplines of the social sciences.

The next chapter reviews the general methodologies and specific procedures used in developing the materials. The materials were used to test the questions raised in discussions from this and earlier chapters.

Chapter III

Development of Materials Used in the Study

In the present study, the development of the treatment materials, the treatment procedure, and the testing instruments were of primary importance. This chapter describes the development of four elements: 1) curriculum materials; 2) treatment procedures; 3) characteristics and construction of the geography achievement test; and 4) alterations in treatment procedures.

Construction of the Curriculum Materials

Treatment preparation for the experiment consisted of the development of the unit Functions of Cities (Jones, 1974). The same student text was developed for both the mastery (T_1) and the non-mastery (T_2) learning groups. The mastery workbook differed in amount of correction and feedback; content and workbook exercises were identical in the mastery and non-mastery workbooks.

Text Content

The text Functions of Cities consisted of nine chapters, as listed in Table 3.1. Chapter 1 "Economic Base and Function" introduced the two main generalizations, "function" and "economic base." Function was defined and illustrated in terms of the relation of the city to the economy of the country; economic base was defined and illustrated in terms of the way people in a city depended upon the most important economic activities for their livelihood. The introductory chapter also gave an overview of the eight cities and their functions and was designed to serve as an advance organizer to the unit (Ausubel, 1963). The next eight chapters then gave a descriptive and analytical presentation of eight cities in terms of their salient economic functions. The text concluded with a glossary.

Table 3.1

Cities by Geographic Location and Function,

Text. Functions of Cities, Jones, 1974

Table of Contents

Chapter	City	Country	Continent	Function
1	Introduction			
2	Durban	South Africa	Africa	Port
3	Frankfurt	West Germany	Europe	Commerce
4	Pittsburgh	United States	North America	Industry
5	Brasilia	Brazil	South America	Government
6	Surfers Paradise	Australia	Australia	Resort
7	Benares	India	Asia	Religion
8	Mexico City	Mexico	South America	Dominant City
9	Tokyo	Japan	Asia	Super City

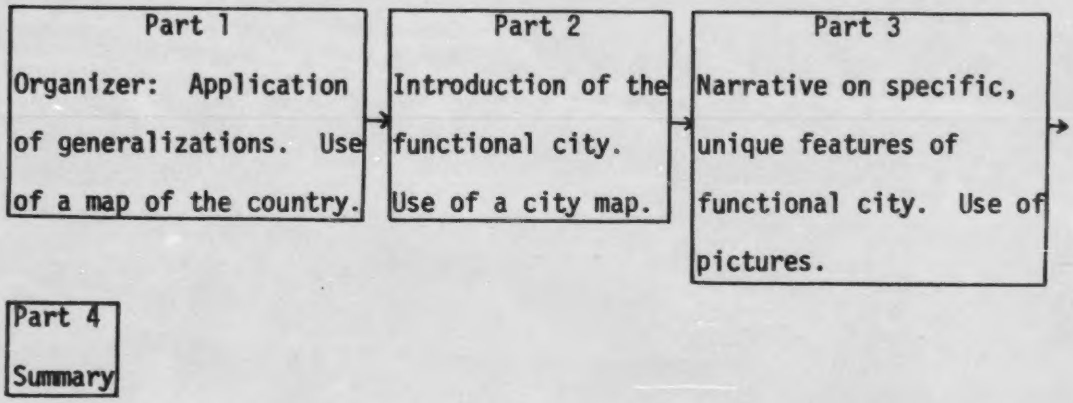
The eight cities were selected to provide type illustrations of function and to give geographic coverage of all the continents, except Antarctica. Europe and North America were underrepresented in terms of geographic coverage. However, each city served as an example of the function of a city with certain economic characteristics. Thus, while Durban was selected as an example of a port city, other characteristic port cities, such as New Orleans, Rotterdam, or Fremantle, might have been selected.

Other considerations than type criteria entered into city selection. The recency of the development of Brazilia as well as its modern planning and buildings were influential in the selection of this city as a government type. Mexico City was selected as the dominant city not only because of its commanding position within the economic and political life in Mexico, but also because of the interest the Geography Curriculum Project has in Mexico as a potential area of field study in connection with the new Latin American Studies Program in the Department of Social Science Education. Tokyo was selected as an example of a super city not only because of its function in the world economy, but because it serves as an example of a non-western city achieving international prominence.

Categorization of cities by a particular function was not used as a device to restrict discussion of the interrelationship of economic activities. Each chapter attempted to show that while a city might be categorized by a function, with a principal economic base, economic activities interact. The writer believed that this method of presentation not only had the merit of contributing geographic diversity to the presentation, but also permitted an intensive development of the conceptual economic base that relates to the modern urban environment. The workbook required the student in most cases to apply the knowledge of one type of city to another similar city which has not been studied. Therefore, the text and workbook together provided a basis for a clearer understanding of world urban economics.

Chapter Format

The format for the eight city chapters followed a structured presentation. The basic format was:


Part 1
Organizer: Application of generalizations. Use of a map of the country.
Part 2
Introduction of the functional city. Use of a city map.
Part 3
Narrative on specific, unique features of functional city. Use of pictures.
Part 4
Summary

Part 1. This section acted as an advanced organizer for the specific concepts and facts that followed in Parts 2 and 3. The two major generalizations 'function' and 'economic base' were used within the context of the functional city. A map of the country, locating the city, was used.

Part 2. Each selected city was discussed in general terms to provide an overview of the city and to identify specific pertinent characteristics of the city. A map of the city, locating many of the specific pertinent characteristics, was provided.

Part 3. Each characteristic was developed to provide the student with an examination of the economic forces within the city type. Pictures were used to supplement the narrative.

Part 4. This section provided a succinct summary statement concerning the narrative of the study. Appendix A contains a sample copy of the student text.

The previous sections have described the content of the textbook for the treatment unit. The next section will discuss the workbook content and format.

Workbook Content

The content in the workbooks was the same for both treatment groups. Each chapter in the student text had a parrallel chapter in the workbook. The various activities that the students were required to complete were premised upon the reading and study of material appearing in each chapter of the student text. The learning outcomes expected from each set of workbook exercises was dependent upon the type of activity that had to be completed. Activities within each chapter of the workbook ranged from recall of knowledge and facts to generalizations and applications of concepts. These activities were written at the knowledge and application levels of the Taxonomy of Educational Objectives: Handbook I. Cognitive Domain (Bloom, Englehart, Furst, and Hill; 1956). Questions also focused upon the maps and pictures appearing in the student text. Consequently, students were expected to display observation and map reading skills as well as stated learning skills.

Activities in the workbook were presented in a variety of forms.

Workbook Format

The workbook aided the student to learn new knowledge about the functions of cities. It also provided the student with practice in using the knowledge learned. Practice was provided through the activities that were available in each of the chapters. Each chapter contained a combination of the following activities:

1. Main Words
2. I can match words with definitions.
3. I can write a definition for each main word.
4. I can match an example or illustration of the main words.

5. I can write an example or an illustration of the main words.
6. I can do or explain activities.
7. Thought Questions
8. Review Test(s)

The consistent use of this format provided students with a greater opportunity to operate with the content and knowledge required by each of the activities. Less time had to be spent by the students to decipher how they were to perform the learning task within the scope of the treatment procedures. The only difference between the workbook for the mastery groups was the inclusion of an extra review test. This difference is discussed in the next section.

Treatment Procedures

Two treatment procedures were employed in this study. There was a mastery (T_1) and a non-mastery (T_2) learning procedure. As the materials used in this study were the same for both treatment groups, the focus of the study was on the manipulation of various components within the mastery treatment procedure. The two treatment procedures were conceptualized in the following format:

	Treatment 1 (Mastery)	Treatment 2 (Non-Mastery)
Presentation		
Narrative	X	X
Student Workbook Activities	X	X
Diagnosis		
Review Test One	X	X
Correction	X	0
Feedback	X	0
Remediation		
Prescriptive Review	X	0
Specific Practice	X	0
General Review	X	0
Diagnosis		
Review Test Two	X	0
Correction	X	0
Summative Test		
(Administered to all students at the conclusion of the unit)	X	X
Weekly Class Discussion	X	X

The X's indicate the components that were used in the procedure while the O's indicate those components not used.

In order to learn material in the text, the workbook provided mastery and non-mastery procedures.

Procedures Common to Both Treatments

All students regardless of treatment had to follow steps 1-9: Appendix B is a sample copy of the non-mastery (T_2) workbook.

1. Fill in the time log at the beginning of each chapter with time work begun and the date.

2. Read one chapter of the student text. Begin with Chapter One.

3. When ready open the workbook, close the text, and work through the activities.

4. When the activities were completed the student turned to the answer sheets at the back of the workbook and corrected his work.

5. If any activities were incorrect the student re-read the text and then did the activities over.

6. When the student was ready the student indicated to the teacher readiness to take a review test.

7. The review test was self-administered.

8. The student corrected the review test from the answer sheets at the back of the workbook. As each review test contained 20 items a score out of 20 was recorded.

9. The classroom teacher checked the results of the test and non-mastery students completed the time log with the ending time and the number of minutes worked. Non-mastery (T_2) learning students the proceeded to the next chapter and followed the same procedures. However, mastery (T_1) learning students were required to perform remedial learning tasks.

Additional Procedures for Mastery Learning

Procedures 10-20 applied only to students in the mastery treatment: (See Appendix C for a sample copy of the mastery workbook).

10. The criterion level arbitrarily selected for this study, but supported by the studies of Block (1970) and Kim (1969), was 85 per cent. This meant that on a review test of 20 items students needed 17 items correct to reach the minimum criterion level. If a student got 17 out of 20 items correct or better the student proceeded to the next chapter and repeated the same procedures.

11. If a student got less than 17 items correct out of 20 the student looked at the incorrect items on the test. Each item contained a key beside it. e.g. 1.2A. The 1.2 refers to chapter one, page 2 in the student text, and the A refers to the specific paragraph on that page. This paragraph contained the correct answer to the question.

12. Students were directed to re-read the paragraph in the text for the incorrect test item.

13. Students were directed to correct incorrect workbook items.

14. When all incorrect items had been corrected, the student was directed to review all the work in both text and workbook.

15. Students then informed their teacher of their readiness to take a second review test.

16. Students then self-administered the second review test.

17. Students corrected the second review test from the answer sheets at the back of the workbook.

18. The classroom teacher was then presented with the completed second review test, who recorded whether mastery had been reached or not.

19. Students then completed the time log for the chapter.

20. Mastery students, then proceeded to the next chapter, whether or not they had achieved mastery.

The above workbook procedures were designed so that the student could study and learn without direct instruction by the teacher. Therefore, students were responsible for the learning tasks, correction to find errors, and the remedial stages of the learning procedure.

The weekly class discussion deserves special comment. During the initial administration of the materials to the students, teachers were instructed to provide students with a break from the self-instructional mode every three days. Both the researcher and his major porfessor felt that this might alleviate such problems as boredom and work fatigue that appeared in other self-instructional studies (Dumbleton, 1973; Pelletti, 1973). After a week of instruction with the materials, however, the three-day discussion was waived in favour of having a weekly class discussion on Wednesday during the middle of the school week. All teachers reported that this was a more satisfactory arrangement. Both students and teachers reported at the conclusion of treatment that the weekly class discussion was a major contribution to maintain student interest and perseverance.

Readability

Functions of Cities was written for students in the middle grades. The most appropriate way to determine whether the materials were satisfactory for this age level would have been to administer them across a broad cross-section of levels. However, the limited resources of the researcher precluded the use of this approach and consequently did not allow a complete evaluation of the materials. A second method is to establish readability by using a standardized reading formula. The three most commonly used readability formulas are Dale-Chall, Spache, and Flesch. In this study the Rudolf Flesch (1949) formula for readability was applied. It was selected primarily because it is more appropriate for materials used with upper Elementary and Junior High School materials. Second, the Flesch formula does not rely upon a specific word list (e.g. Dale-Chall) which can get out of date (Powers, Sumner, and Kearl, 1958).

The Flesch formula requires that a number of steps be followed. First, the number of words must be counted per sentence.

(a) Count as a sentence each unit of thought that is grammatically independent of another sentence or clause. Its end may be marked by a period, question mark, exclamation point, semi-colon, or colon. Also count a fragment as a sentence.

(b) To do this, count the words in ten sentences separately, add, then divide by ten.

Second, count the syllables in 100 words. When these tasks have been completed. The following arithmetic operations should be conducted:

(a) Multiply the average sentence by 1.015

(b) Multiply the number of syllables in 100 words by .846

(c) Add (a) and (b)

(d) Subtract this sum from 206.835

(e) This provides the Reading Ease Score

The Reading Ease Score is then tested against a table which provides information concerning the readability level of materials. In order to obtain a readability level for Functions of Cities, two samples were selected from Chapters 2-9. Table 3.2 shows the Reading Ease Scores obtained for the samples.

The mean of the Reading Ease Scores was 85. When this score was tested against the readability scale it was found that the materials had a readability level of Grade Six.

Construction and Characteristics of the Review and Summative Achievement Tests

This section describes the construction and characteristics of the review tests and the summative test (Geography Achievement Test), along with the methods used to establish validity and reliability. Nine review tests and one summative test were constructed.

Construction of the Review Tests

Each chapter in the workbook Functions of Cities contained a review test. A review test measured the amount of learning in a chapter. Each review test contained 20 items. Items were written in three forms: 1) three foil, multiple choice; 2) true or false; and 3) completion. There was no consistent number of items in each form. Some chapters contained more multiple choice items while others contained more true or false, or completion items.

All review items were written strictly on the content in the student text. Each test item was keyed to a particular paragraph within the chapter. Items tested recall, application, and transfer of cognitive knowledge.

The non-mastery (T_2) treatment contained one review test at the conclusion of each chapter. Review tests used in both treatments were exactly the same. The mastery (T_1) treatment contained two review

Table 3.2

Reading Ease Scores and Grade Level for 17 Samples Selected from Nine Chapters of the Materials Functions of Cities

Chapter	Sample	Reading Ease Scores	Grade Level
1	1	90	5
2	2	93	5
	3	97	5
3	4	80	6
	5	85	6
4	6	89	6
	7	83	6
5	8	63	8 or 9
	9	93	5
6	10	93	5
	11	89	6
7	12	67	8 or 9
	13	100	4
8	14	93	5
	15	86	6
9	16	68	8 or 9
	17	73	7

tests. The second review test for the mastery (T_1) treatment did not contain new test items; the items used in the first review test were merely reordered. The review tests can be seen in the workbooks in Appendices B and C.

Content Validity

Items were constructed on the premise that the text contained the knowledge necessary to answer the questions. The researcher constructed the items then keyed them to a particular paragraph in the chapter. The researcher's major professor Dr. Marion J. Rice and graduate students James S. Fagan and Robert R. Myers then checked the text and the test items to establish that the test items measured the knowledge conveyed in the text. Changes were made at their suggestion to reduce ambiguities, improve form, and simplify language.

Knowledge and application items were constructed in accordance with the Taxonomy of Educational Objectives: Cognitive Domain (Bloom, et. al., 1956) to measure student learning on each of the review tests.

Reliability

Within the context of the mastery learning procedure, the concept of reliability applied to the review tests was considered inappropriate. The review tests were criterion referenced tests, not norm referenced tests. A criterion referenced test requires that students perform to an arbitrarily selected criterion level. This study used the 85 per cent achievement level on a review test as the criterion level. A norm reference test is used to determine an achievement score for individual students. The scores vary from student to student. The scores have a range. Because students were expected to achieve to a criterion level, and hence there was little score variance, no reliability measures were obtained for the review tests (Gronlund, 1973; Popham and Husek, 1971). However, the summative test was treated differently.

Construction of the Summative Test

The final version of the summative test was in two parts. The first part was a 40 item, four option, multiple choice test while the second part was a 24 item, retrieval chart completion test. The total

64 item test was designed to measure the students' knowledge of facts, concepts, and generalizations presented in the treatment unit. The procedures followed in constructing this test are outlined below:

1. The major facts, concepts, and generalizations to be learned were identified in the treatment unit. (See Appendix D, p.462 for a list.)

2. A table of specifications was drawn up and each content item was categorized for inclusion in the table. (See Appendix E, p.464 for table of specifications.)

3. A 40 item, four option, multiple choice was constructed. This task was simplified because items for the major facts, concepts, and generalizations had been previously constructed for the review tests. However, an extra option was added to each of the items selected and in many cases the items were rewritten and reworded. A 24 item retrieval chart was also constructed. A clue was provided for each of the eight cities that were studied in the treatment unit and three extra pieces of information were needed to fill in the blanks. (See Appendix F, p.467 for the 64 item test.)

4. Dr. Marion J. Rice and the researcher were solely responsible for the writing and selection of the final summative test. No other people were as familiar with the content of the unit and the student learning outcomes that were expected. The items selected were not only appropriate to the content, but also appeared to display clarity, understandability, and accuracy.

Content Validity

Due to the process described above it was believed, by the researcher, that no rules associated with content validity had been violated. Consequently, it was assumed that the summative test met the criteria for content validity.

Reliability

The summative test was a norm referenced test. Students responded to items to the best of their knowledge and the scores from the summative test were used as data for purposes of statistical analysis. Unfortunately, due to the press of time no pilot testing of the measuring instrument was conducted. Instead the following procedures were followed:

1. The 64 item summative test was administered to the treatment groups.

2. An arbitrary decision was reached by the researcher to select from one of the treatment groups three classes that displayed the widest range of scores as measured by the class mean on the measure used as the blocking variable the Iowa Tests of Basic Skills word meaning section.

3. The class scoring the highest mean and the class scoring the lowest mean both fell in the mastery treatment group. Both these classes were selected. Another class in the mastery treatment group closest to the mean was also selected. In all 76 students were used.

4. Student responses on the 40 item, four option, multiple choice test were transposed to IBM sheets. These scores were then analyzed by the Analysis of Item and Test Homogeneity (ANLITH) computer program. Table 3.3 summarizes the ANLITH results.

Table 3.3

Test Analysis Data for the 40 Item Multiple Choice Instrument

Grade	Number of Students	Number of Questions	Estimate of Reliability	Mean	S.D.	S.E. of Measurement
7 (Three Classes)	76	40	.89	21.40	8.34	2.76

5. The results of the ANLITH indicated that the 40 item multiple choice test had a reliability of .89.

6. Item difficulty was examined for each item on the test. Two items (number 12, 36) had high difficulty (under 30 per cent scored correctly) while four items (numbers 16, 18, 26, 28) had low difficulty (over 70 per cent scored correctly). As the test had already been administered no changes were made. (See Appendix F, p.467 for the 40 item test.)

Table 3.4

Test Analysis Data for the 24 Item Recall Instrument

Grade	Number of Students	Number of Questions	Estimate of Reliability	Mean	S.D.	S.E. of Measurement
7 (Three Classes)	69	24	.95	13.00	7.69	1.77

7. A reliability analysis was conducted with the 24 item recall test. The results of the ANLITH indicated that the 24 item recall test had a reliability of .95.

8. As the test had already been administered no changes were made. However, the low standard error of measurement of 1.77 was an indication that this was a test that could measure individual's knowledge of recall. (See Appendix F, p.467 , for the 24 item recall test.). Tables 3.5 and 3.6 show the mean scores and percentages of the 40 item and 24 item posttest measure by treatment and aptitude. These tables show that there remained a range of scores consistent with the aptitude levels of the students used in the present study. High aptitude students achieved higher than middle and low aptitude students, as did middle aptitude students achieve higher than low aptitude students.

Alterations in the Treatment Procedures

Most experimental studies that use classroom learning materials conduct a pilot test of the materials and the treatment procedures. This study did not employ a pilot test phase because the time remaining in the school year after material development and duplication did not allow for a pilot run. It was therefore necessary to test the mastery procedure without a pilot trial. Detailed and careful procedures described previously had been derived for the mastery and non-mastery treatments.

During the first week of instruction teachers were requested to monitor the treatment procedures carefully and to observe the reaction of students. A Report From Teachers form was provided to each teacher. See Appendix G, p.479 for a copy of the Report from Teachers.) Questions concerning the treatment procedures, content, and student and teacher reactions were included on the report. Teachers indicated that

Table 3.5

Mean Scores for the 40 Item, 24 Item and Total Achievement Measure by Treatment and Aptitude

	Class	1			2			3			4			5			6			7			8			9			10		
		MC[1]	R[2]	T[3]	MC	R	T	MC	R	T	MC	R	T	MC	R	T	MC	R	T	MC	R	T	MC	R	T	MC	R	T	MC	R	T
MASTERY	HIGH	23	18	41	14	6	20	26	18	44	26	13	39	30	19	49	28	19	47	24	11	35	24	16	40	24	15	39	18	10	28
	MIDDLE	17	13	30	19	5	24	21	13	34	25	14	39	24	18	42	20	12	32	14	3	17	27	16	43	17	6	23	16	2	18
	LOW	15	5	20	14	5	19	10	7	17	13	3	16	7	3	10	12	3	15	14	2	16	13	2	15	12	0	12	12	1	12
	Class	11			12			13			14			15			16			17			18			19			20		
NON MASTERY	HIGH	29	20	49	22	9	31	22	4	26	27	17	44	30	17	47	20	10	30	20	13	33	24	14	38	23	7	30	12	3	15
	MIDDLE	19	18	37	19	5	24	14	3	17	20	12	32	16	11	27	14	2	16	16	1	17	16	9	25	22	6	28	15	1	16
	LOW	14	5	19	15	2	17	11	1	12	20	11	31	11	10	21	18	2	20	13	2	15	12	4	16	8	2	10	11	1	12

1. M.C. means Multiple Choice and refers to the 40 item test.
2. R. means Recall and refers to the 24 item test.
3. T.means Total and refers to the combined scores on the posttest.

Table 3.6

Percentage Correct for the 40 Item and 24 Item Achievement Measure by Treatment and Aptitude

	Class	1		2		3		4		5		6		7		8		9		10	
MASTERY		MC[1]	R[2]	MC	R	MC	R	MC	R	MC	R	MC	R	MC	R	MC	R	MC	R	MC	R
	HIGH	58	75	35	25	65	75	65	54	75	79	70	79	60	46	60	67	60	63	45	42
	MIDDLE	43	54	48	21	53	54	63	58	60	75	50	50	35	13	68	67	43	25	40	8
	LOW	38	21	35	21	25	29	33	13	18	13	30	13	35	8	33	8	30	0	30	4
	Class	11		12		13		14		15		16		17		18		19		20	
NON MASTERY	HIGH	73	83	55	38	55	17	68	71	75	71	50	42	50	54	60	58	58	29	30	13
	MIDDLE	48	75	48	21	35	13	50	50	40	46	35	8	40	4	40	38	55	25	38	4
	LOW	35	21	38	8	28	4	50	46	28	42	45	8	33	8	30	17	20	8	28	4

1. M.C. means Multiple Choice and refers to the 40 item test.
2. R. means Recall and refers to the 24 item test.

the procedures were working satisfactorily. However, they did indicate that a change from a three day break to a mid-week break might improve student learning and maintain interest. This change was instituted. The same report was given to the teachers after each succeeding week of instruction. All reports were satisfactory.

The content of the text and the workbook had already been set. Consequently, even if satisfaction had not been expressed by the teachers there was very little that could have been done to alter the content. All teachers, however, expressed concern for the reading level of the materials. It should be noted that the mean reading level of the population of students used in this study was 6.6 while the reading level of the materials was Grade Six. The researcher decided that this criticism was not germane in the light of these statistics.

A pilot test is an important part of an experimental study from the standpoint of both test construction and treatment procedures. However, it was not possible to conduct such a pilot test study. While certain precautions were built into the actual administration the researcher acknowledges that the absence of pilot testing is a limitation of the study.

Summary

This chapter outlined the development of the curriculum materials and measuring instruments used in the study. The curriculum materials described the economic base and function of selected cities around the world. The treatment units used in the study were constructed in two formats. Treatment 1 was a mastery learning procedure while treatment 2 was non-mastery learning procedure.

The remainder of the chapter described the format and construction of the review tests and the summative geography achievement test. Finally, an explanation was offered for the omission of the pilot test. However, a number of precautions were described that should have offset the disadvantages of the lack of the pilot test. The researcher acknowledged that this was a limitation to the study.

The next chapter will describe the research design and the statistical procedures used to analyze the data.

CHAPTER IV

METHODOLOGIES AND PROCEDURES

This chapter describes the following six elements of the study: 1) experimental design; 2) experimental study; 3) pattern of logic used in the study; 4) contextual variables; 5) statistical procedures; and 6) limitations.

EXPERIMENTAL DESIGN

A 3 x 10 x 2, aptitude by classes-nested-within-treatments, by treatments, multivariate analysis of variance (MANOVA) using three measures of effect was employed with the posttest data of this study. This design is shown in Table 4.1.

Rationale for the Design

This design was used in order to counter the main disadvantage of completely randomized designs--their relative inefficiency. The error term, against which the variability of treatment means is tested, is generally large in randomized designs. This large error term results from the variability among subjects within groups. Much of the error variance arises from individual differences in factors which effect performance. The blocking design is one method of removing some of the error variance due to individual differences (Myers, 1966).

Blocking has four advantages. First, the treatment groups are roughly matched on a measure which should affect performance. Second,

Table 4.1

Experimental Layout (MANOVA)

	Treatment 1 - Mastery				
	Class 1	Class 2	Class 3		Class 10
R_1 Upper Aptitude Group	$\bar{Y}_{\underline{1}111}$. $\bar{Y}_{\underline{1}113}$	$\bar{Y}_{\underline{2}111}$. $\bar{Y}_{\underline{2}113}$	$\bar{Y}_{\underline{3}111}$. $\bar{Y}_{\underline{3}113}$	.....	$\bar{Y}_{\underline{10}111}$. $\bar{Y}_{\underline{10}113}$
R_2 Middle Aptitude Group	$\bar{Y}_{\underline{1}211}$. $\bar{Y}_{\underline{1}213}$	$\bar{Y}_{\underline{2}211}$. $\bar{Y}_{\underline{2}213}$	$\bar{Y}_{\underline{3}211}$. $\bar{Y}_{\underline{3}213}$	.....	$\bar{Y}_{\underline{10}211}$. $\bar{Y}_{\underline{10}213}$
R_3 Lower Aptitude Group	$\bar{Y}_{\underline{1}313}$. $\bar{Y}_{\underline{1}313}$	$\bar{Y}_{\underline{2}311}$. $\bar{Y}_{\underline{2}313}$	$\bar{Y}_{\underline{3}311}$. $\bar{Y}_{\underline{3}313}$	.....	$\bar{Y}_{\underline{10}311}$. $\bar{Y}_{\underline{10}313}$

	Treatment 2 - Non-Mastery				
	Class 11	Class 12	Class 13		Class 20
R_1 Upper Aptitude Group	$\bar{Y}_{\underline{11}121}$. $\bar{Y}_{\underline{11}123}$	$\bar{Y}_{\underline{12}121}$. $\bar{Y}_{\underline{12}123}$	$\bar{Y}_{\underline{13}121}$. $\bar{Y}_{\underline{13}123}$	.....	$\bar{Y}_{\underline{20}121}$. $\bar{Y}_{\underline{20}123}$
R_2 Middle Aptitude Group	$\bar{Y}_{\underline{11}221}$. $\bar{Y}_{\underline{11}223}$	$\bar{Y}_{\underline{12}221}$. $\bar{Y}_{\underline{12}223}$	$\bar{Y}_{\underline{13}221}$. $\bar{Y}_{\underline{13}223}$	.....	$\bar{Y}_{\underline{20}221}$. $\bar{Y}_{\underline{20}223}$
R_3 Lower Aptitude Group	$\bar{Y}_{\underline{11}321}$. $\bar{Y}_{\underline{11}323}$	$\bar{Y}_{\underline{12}321}$. $\bar{Y}_{\underline{12}323}$	$\bar{Y}_{\underline{13}321}$. $\bar{Y}_{\underline{13}323}$	.....	$\bar{Y}_{\underline{20}321}$. $\bar{Y}_{\underline{20}323}$

R represents reading (word meaning) levels: $r = 3$

R_1 is the upper reading level. R_2 is the middle reading level. R_3 is the lower reading level.

T represents treatments: $t = 2$

T_1 received the mastery treatment and T_2 received the non-mastery treatment.

C represents classes: $c = 20$

Subscript Order: class, aptitude, treatment, measure the interaction effects can be studied. Third, the blocking design will usually be more efficient than a "one factor design involving the same total number of dependent measures at each treatment level (Myers, 1966). Four, the blocking design allowed the researcher to observe the efficiency of the mastery procedures with students of varying aptitude, particularly low aptitude students. However, it was not a practical possibility to rearrange student seating into aptitude levels in each of the classes to minimize across aptitude level interaction. Students within each class sat at their normal work desk. No attempt was made by the researcher to control for across-aisle or within-aisle student communication even though the self-instruction materials were designed to minimize student interaction.

This design also involved the use of two posttest treatment groups. There were several reasons why a posttest-only, rather than a pretest-posttest design was used. As Campbell and Stanley (1963) have pointed out, the pretest of initial differences is not essential in experimental designs. The randomization of students to the two treatment groups controlled for initial systematic biases. Since randomization controlled for systematic initial biases, it was assumed that the achievement scores of the two treatment groups would have exhibited only chance differences from each other on a pretest.

A cognitive pretest was also rejected. Campbell and Stanley (1963) have indicated that a pretest of new subject matter is inappropriate. Greene (1965), Thomas (1967), and Walsh (1967) found that pretest scores of students did not differ significantly from chance. These findings suggested that pupil scores on a pretest in the present study probably would not have differed significantly from chance.

According to Campbell and Stanley (1963), the Posttest-Only Design is preferred to the Pretest-Posttest Design because it controls for the effects of the pretests. Pretesting may have been a confounding variable in the proposed study. The Posttest-Only Design also required only two treatment groups, thereby resulting in larger sample sizes than would have been possible if other research designs had been selected which required more than two treatment groups, such as the Solomon Four Group Design.

As a practical matter, moreover, teachers and students react negatively to the administration of pretests on subject matter with which they have had no systematic instruction. Informal observations with studies using pretests (Greene, 1965; Thomas, 1967; and Walsh, 1967) indicate that the administration of a pretest can lead to a hostile attitude on the part of students to an experimental study. Since the population selected as mastery and non-mastery subjects were not accustomed to using self-instructional materials over long periods of learning, procedural treatment prudence as well as design considerations supported the desirability of a posttest-only design.

Rationale for the Concomitant Variable

In the conduct of experimental research, standardized measures may be used for a variety of purposes--to predict pupil achievement, to match sample to reading level of material, to describe pupil cognitive variables, and to establish concurrent validity of the instrument developed by the investigator. Since 1965, a continuing concern of the Anthropology Curriculum Project and the Geography Curriculum Project has been to develop materials for pupils in terms of characteristics related to school achievement. Reading ability has been consistently identified as the most significant ability related to success in school. As a practical matter, the typical full scale reading battery, such as in the Iowa Tests, takes more than one class period to administer, a practical matter which interferes with collection of data. The word meaning, or vocabulary sections, of most reading or achievement tests can easily be administered within the time constraints of one period within a classroom. Because of the high correlation of vocabulary to reading, the Anthropology and Geography Curriculum Projects have therefore used knowledge of word meaning, as measured by a vocabulary test, as an efficient way to collect data for the concomitant variable.

The concomitant variable selected for this study was knowledge of word meaning, as measured by the vocabulary section of the _Iowa Tests of Basic Skills: Forms 5 and 6_ (Lindquist and Hieronymus, 1971). Administration time is 17 minutes.

Knowledge of word meaning was selected as the concomitant variable because this category correlates highly both with the ability to read and to achieve in school subjects. Russell (1961) writes that many well known standardized reading tests, including the _Iowa Tests of Basic Skills_, contain tests of vocabulary meaning. A child's understanding and interpretation of sentences and paragraphs will depend considerably upon his knowledge of individual words in the larger units. The Iowa vocabulary test was used as the concomitant variable in the present study. Knowledge of word meaning correlates more highly with reading comprehension than any other sub-test of the Iowa battery (Technical

Manual, 1956, 1964, 1971).

The reading test of the Iowa battery is a reading comprehension test (Morgan, 1959). Knowledge of word meaning is essential to the ability to read, and is widely used in the testing of students to predict subsequent success in school (Seegers, 1939; Spache, 1943; Traxler, 1945).

Knowledge of word meaning is also the subtest on the Binet and Weschler that consistently show the highest correlation with the total score (Thorndike and Hagen, 1969) and, the first sub-test on the Binet, which is used to establish difficulty of testing level. This high correlation of success in school with verbal ability has stimulated the development of picture-vocabulary tests as abbreviated intelligence test devices. Examples of these picture-vocabulary tests are the Full-Range Vocabulary Test (Ammons and Ammons, 1948) and the Peabody Picture Vocabulary Test (Dunn, 1959).

The word meaning section of the Iowa Test was also chosen for high test reliability and use in Georgia state-wide testing. According to the 1974 Technical Manual, the grade seven vocabulary test obtained a test reliability of .89, while the reading test obtained a reliability of .92. The intercorrelation between the vocabulary and reading test was .81. The standard error of measurement on the raw scores for the vocabulary test was 3.0. With the large sample used in obtaining the reliability data this is a strong indication that the vocabulary test was predicting vocabulary level highly.

The Iowa Test battery is used in the Georgia state-wide testing program at Grades 4, 8, and 12. Consequently, the use of the Iowa Test is readily accepted in curriculum research by Georgia teachers and administrators.

Unit of Statistical Analysis

The researcher, in this study, had three choices when a unit of statistical analysis was chosen. The choices were the individual, the classroom, or the aptitude group. The individual was not the focus in this study even though the treatment materials were self-instructional. The classroom should have been the unit of statistical analysis; however, this would not have allowed an analysis of the relationship between the three aptitude levels and treatments. Therefore, the aptitude group was used as the unit of statistical analysis. The procedure for randomization and cell assignment is described in the section "Random Assignment."

Experimental Study

This study compared self-instructional mastery and non-mastery treatments to determine if there were differences in achievement and time of high, middle, and low aptitude students on learning, retention, and times-to-testing.

Sample Selection

Dr. Marion J. Rice, Director of the Georgia Geography Curriculum Project, made arrangements with officials of the Savannah-Chatham County Public Schools in Georgia to obtain 20 Grade Seven classes (539 students) in four schools for the experimental study.

Random Assignment of Individuals to Treatment Groups

There were five steps in the randomization process. First, all students were administered the word meaning section of the *Iowa Tests*

Table 4.2

Number of Students by Reading Level by Class and Treatment Including those Students Omitted from Data Analysis

Treatment by Aptitude		Classroom 1	2	3	4	5	6	7	8	9	10
MASTERY	Upper	5	2 $\boxed{1}$	10 (2) $\boxed{8}$	6 $\boxed{2}$	22 (2) $\boxed{2}$	5 $\boxed{3}$	6 $\boxed{3}$	9 $\boxed{3}$	3 (2)	6 (1) $\boxed{1}$
	Middle	4 $\boxed{2}$	3 $\boxed{1}$	3	5 (1) $\boxed{2}$	3	9 $\boxed{3}$	6 $\boxed{2}$	3 $\boxed{4}$	8 $\boxed{1}$	7 (1) $\boxed{6}$
	Lower	12 (3)	13 (3)	2	13	1	8 (1)	9 (3)	10 (1)	9 (2)	6 (1)

Treatment by Aptitude		Classroom 11	12	13	14	15	16	17	18	19	20
NON MASTERY	Upper	5 $\boxed{1}$	4 $\boxed{1}$	1 (1)	14 $\boxed{4}$	14 (1) $\boxed{4}$	6 $\boxed{1}$	5 $\boxed{1}$	5 (1)	9 $\boxed{2}$	2 $\boxed{1}$
	Middle	3 (2) $\boxed{6}$	5 (1) $\boxed{4}$	8 $\boxed{3}$	5 $\boxed{2}$	6	8 $\boxed{4}$	5 $\boxed{9}$	5 $\boxed{8}$	3 $\boxed{3}$	2 $\boxed{3}$
	Lower	12	8 (2)	11 (2)	2 (1)	1 (2)	4 (3)	9	10	8	10 (1)

The bracket ( ) indicates the number of students dropped from the study.

The square ☐ indicates the number of students not used for data analysis.

of Basic Skills: Forms 5 and 6 (Lindquist and Hieronymus, 1971). Student scores were rank ordered and a mean and standard deviation was computed for the group. Second an a priori decision was made to select reading aptitude groups within classes based upon the mean of student scores, creating cells of unequal Ns. The mean of the group was 19.03 and a quarter of a standard deviation on either side of the mean formed the middle reading aptitude group. The gaps between one quarter and one half standard deviations above and below the mean were used as clear differentials between the three levels of reading aptitude. Students falling within these deviations participated in the study but they were excluded in the data analysis. The high reading aptitude group was comprised of students whose scores were greater than one half standard deviation above the mean. The low reading aptitude group was comprised of students whose scores were more than one half standard deviation below the mean. Third, students were then sorted back into their classes maintaining their respective aptitude grouping. Fourth, classes were then randomly assigned to one of two groups. Fifth, treatment was then randomly assigned to the groups.

Distribution of Students by Treatment and Aptitude

Twenty grade seven classes (539 students) were selected for this study. Students within classes were distributed as displayed in Table 4.2. All 539 students were not used in the study. There were two basic reasons why some students were not used. First, when the treatment by levels was set up on the concomitant variable some student scores on the word meaning test fell into the groups between aptitude levels. This occurred because this study required a clear differentiation between the high, middle, and low aptitude groups. This resulted in 19 per cent of the population being omitted from the data analysis.

Second, students were deliberately omitted from the analysis for the following reasons:

1. Consistent and prolonged absences from school for more than ten of the twenty instructional days. Absences due to sickness or suspension were the only explanations excepted in this category.

2. Students had moved away from the school and either did not complete the unit materials or could not complete the final tests.

This resulted in 7 per cent of the population being omitted from the data analysis of which 26 per cent came from the high aptitude group, 10 per cent came from the middle aptitude group, and 64 per cent came from the low aptitude group.

Reading Scores, Grade Equivalents, and National Percentile Rank

Students distributed by aptitude contained the following characteristics. High aptitude students were reading equivalent to grade level. Middle aptitude students were approximately two grade levels lower, while low aptitude students were four grade levels below actual grade level (see Table 43 and 4.4 for aptitude and grade equivalent levels. The grade equivalent scores translated to national percentile ranks indicate that the high aptitude group fell in the 58th percentile rank, the middle aptitude group fell in the 25th percentile rank, and the low aptitude group fell in the 3rd percentile rank (see Table 4.5). These scores indicate that most students used in this study were below the national norm for reading as measured by the word meaning section, Iowa Tests of Basic Skills.

Table 4.3

Mean Reading Scores, Standard Deviations, and Grade Equivalents by Treatment and Aptitude Level

	Class	1	2	3	4	5	6	7	8	9	10	Total	G.E.*
MASTERY	High	$\bar{x}$ 25.0 s 0.0	24.5 0.0	26.7 2.4	25.7 1.2	33.3 5.3	32.6 6.6	28.0 3.4	29.4 3.9	30.6 5.6	30.1 4.7	29.1	8.1
	Middle	18.5 1.3	19.3 2.1	19.0 1.0	19.7 1.2	18.0 1.0	19.2 1.0	18.8 1.0	20.7 .7	19.1 1.7	18.4 1.4	19.1	6.3
	Low	9.4 3.2	10.6 2.6	13.0 1.4	11.3 2.6	12.0 0.0	12.0 1.2	11.3 2.2	11.4 2.8	9.2 3.4	9.9 2.5	11.0	4.0
	Class	11	12	13	14	15	16	17	18	19	20	Total	G.E.*
NON MASTERY	High	28.6 3.2	28.0 3.6	27.0 0.0	30.9 5.0	32.2 7.1	30.0 5.1	29.8 5.1	31.2 4.8	31.6 7.0	27.0 4.2	29.6	8.1
	Middle	18.6 1.3	19.0 1.7	18.5 1.1	19.2 2.1	18.5 1.7	18.5 1.7	19.6 1.7	19.4 1.5	20.3 .7	18.0 1.41	19.0	6.3
	Low	9.8 3.1	10.5 2.4	8.5 3.3	10.3 3.5	8.3 4.9	9.9 3.2	12.1 2.4	12.4 1.6	10.3 2.7	11.2 2.9	10.3	3.8

* Grade Equivalents

$\bar{x}$ equals the mean of the cell.
s equals the standard deviation from the mean of the cell.

Table 4.4

Mean Reading Scores, Standard Deviations, and Grade Equivalents for Aptitude Groups

Aptitude	Mean Scores	S. D.	Grade Equivalents
High	29.4	5.26	8.1
Middle	19.1	2.32	6.3
Low	10.7	2.83	3.9

Table 4.5

Mean Reading Grade Equivalents and National Percentile Ranks for Aptitude Groups

Aptitude	Grade Equivalent	Percentile Rank
High	8.1	58
Middle	6.3	25
Low	3.9	3

Orientation of Teachers

The researcher supplied each teacher and principal of the four cooperating junior high schools copies of the text and workbook Functions of Cities and written instructions regarding procedures. Because the teachers were not required to teach students the treatment unit no attempt was made to train the teachers in any aspect of the treatment material. However, because all learning material was included in the text and workbook teachers were expected to keep abreast of the content, including the Thought Questions. Thought Questions could be used as points for discussion on the mid-week break.

Duration of the Study

The study was conducted over a 20-day instructional period from April 4th to May 7th, 1974. During this period both treatment groups studied Functions of Cities. At the end of the 20-day instructional period a geography achievement posttest was given to both treatment groups. A delayed posttest of geography achievement was administered on May 24th, 1974, 17 days after the conclusion of treatment to measure retention.

Pattern of Logic Used in the Study

A 3 x 10 x 2, aptitude by classes-nested-within-treatments, by treatments, multivariate analysis of variance was used with learning, retention, and times-to-testing as the effects measures. Factors included two treatments and three levels of aptitude. This experimental design was depicted earlier on page 69.

Research Hypotheses

The major purpose of this study was to compare self-instructional mastery and non-mastery treatments to determine if there were differences in achievement and time of high, middle, and low aptitude students. The main hypotheses investigated were:

1. The mastery and non-mastery treatments will produce differences in the average effects which are not the same ($p<.05$) at the high, middle, and low aptitude levels measured by geography posttest of:

(a) learning,

(b) retention

and a measure of,

(c) times-to-testing

2. With pupils pooled across the three levels of aptitude the difference between the master[illegible]on-mastery treatments will produce differences (p<.05) in the average achievement measured by geography posttests of:

(a) learning,

(b) retention

and a measure of

(c) times-to-testing.

3. With pupils pooled across the two treatments, there are differences among the three levels-of-aptitude vectors of average effects (p<.05) measured by geography posttests of:

(a) learning,

(b) retention,

and a measure of

(c) times-to-testing.

Pattern of Logic for Testing the Research Hypothesis Statement

Statement	Logic Pattern	Source
If the research hypothesis is true then the observed differences of average effects will not be the same across the three levels of aptitude.	If A, then B	Assumption

For these differences in average effects to be found different across the three levels of aptitude in the context of the research hypothesis being false is very unlikely.	B without A is extremely unlikely.	Assumption
The differences were found not to be the same across the three levels of aptitude.	A is much more credible	Polya Pattern IV

All research hypotheses followed the same pattern of logic.

Discussion of Pattern of Logic

The pattern of logic used as a base for the proposed study claims that it is extremely unlikely for differences in the average effects to be found different across the three levels of aptitude without the hypothesis being true. This claim can be considered to be probable only if the personal attributes of the subjects and contextual attributes other than treatment are eliminated as possible causes for the differences.

In the proposed study, personal attributes of the subjects can be eliminated as a probable cause of the probability of a Type 1 error $(p<.05)$. This is true because of the randomization factor in the research design. The personal attributes of the subjects other than reading aptitude are randomly distributed along with the assignment of individuals to treatment groups. While randomization does not ensure that the two groups are perfectly matched on all variables which might influence the results of the experiment, it does guard against the danger of systematic biases in the data (Myers, 1966).

The research design does not take into account contextual or situational variables that might cause a difference between group means.

The researcher dealt with these variables in two ways. Whenever possible, direct control of the variables was exercised over such influences as treatment materials, directions to teachers, and test administration.

Where direct control is impractical, variables, e.g. school organization, teacher experience, physical plant or class size, were observed and described systematically.

The direct control of certain contextual variables with the two treatments makes it highly unlikely that those variables caused differences between the means of each treatment group in the study. It was also assumed that variables that were observed and described rather than controlled did not cause a difference in the means of the two treatment groups if the variables did not differ greatly between groups. Within the limits described above, it is logical to claim that any differences in means can probably be attributed to treatment differences, thereby making the assumption more credible.

In the event that the average effects are the same across aptitude levels the claim can still be considered probable due to the control exercised over the subjects and contextual attributes other than treatment which may have accounted for differences in the average effects.

Due to the limitation of experimenting with existing classes which functioned within the framework of the school and the school system, there were some contextual variables that could not be controlled by the researcher. The contextual variables are described in the following section.

Contextual Variables

The contextual variables which could not be controlled included the effects of the community, school district, school, and the teachers.

Community and School District

The study was conducted in the Savannah-Chatham County Public schools. The population of Chatham County is approximately 209,000. The economic base of the city and county is the harbor and docks with the military and manufacturing other important activities.

The student enrollment in the Savannah-Chatham County Public schools was 33,606 as of February 28, 1974. This total systemwide enrollment was composed of 19,292 elementary grade students, 13,353 secondary grade students, 668 elementary special education students, and 143 secondary special education students. There are 17 secondary and 42 elementary schools (B. Hirshberg, personal communication, April 2, 1974).

The school system is under court order to maintain racial balance of faculties and students in every school. This racial balance was achieved by pairing schools with predominant black and white student bodies. Bussing was used to facilitate this equality of racial composition. During the time that this study was conducted, principals and teachers indicated no incidents of racial tension among the students.

Characteristics of the Schools in the Study

The twenty classes that participated in this study were located in four schools in the Savannah-Chatham County School District. These schools contained the following characteristics.

School A. The original construction of the school was completed in 1963. No additions have been contemplated since 1963. There are 24 regular classroom teachers, one special remedial teacher, and one librarian at the school. The school was administered by an appointed principal.

Classes at all grade levels (7-9) were heterogeneously grouped. The racial composition of the school was 54 per cent black and 46 per cent white. Socio-economically, the geographic area around the school was below average and the area was under Title 1 funding. The principal reported that racial tension was not a problem in the school.

School B. The school was constructed in 1960. There were 31 regular classroom teachers and one special teacher. The school was administered by an appointed principal.

The classes were self-contained; however, they were purported to be homogeneous. The word 'homogeneous' was used in the sense that a racial balance was maintained in each class. The racial composition of the school was 50 per cent black and 50 per cent white. Approximately 50 per cent of the school population came from the middle and lower middle class areas around the school, while the other 50 per cent were bussed from economically deprived areas. The principal reported that racial tension was not a problem in the school.

School C. The original construction of the school was completed in 1959. The junior high school is adjacent but integrated with the senior high school next door. There were 28 regular classroom teachers at the junior high school but there were no special teachers. The school was administered by an appointed principal.

The classes were self-contained and heterogeneous. The racial composition of the school was 58 per cent black and 42 per cent white with students coming from the lower, upper lower, and middle socio-economic areas. The principal indicated that the school appeared free from racial tension.

School D. Construction of the school was completed in 1962. There were 46 regular classroom teachers, 2 special teachers, and a librarian. The school was administered by an appointed principal.

Classes were self-contained and heterogenous. The racial composition of the school was 45 per cent black and 55 per cent white. Students came from lower and lower-middle socio-economic areas. The principal did not indicate that racial differences had created any problems.

Characteristics of the Teachers in the Study

Five grade seven teachers from the Savannah-Chatham County School District participated in this study. The researcher spent nine days in Savannah while the study was in progress and during this time considerable observation of classroom and material management was made. The following analysis arose from written teacher responses to a questionnaire and researcher observations.

Teacher A. This teacher was the eldest of the group, female, and had taught for 25 years.

This teacher held a Bachelor of Science degree with a major in social studies. She reported that she had taken nine courses in geography and had attended some geography workshops.

Teacher B. This teacher was in the mid-twenties, male and was teaching for the first time.

He held a Bachelor of Science degree with a major in Physical Education and a minor in social studies. He had completed one course in geography.

Teacher C. This teacher was in the mid-twenties, female, and had four years teaching experience.

She held a Bachelor of Science (Education) degree with a major in Social Science Education in geography. She had completed 55 quarter hours in geography.

Teacher D. This teacher was in the mid-twenties, female and had four years teaching experience.

She held a Bachelor of Science degree in Education with a major in social studies. She had completed one course in geography.

Teacher E. This teacher was in the mid-twenties, male, and had three years teaching experience.

He held a Bachelor of Science degree in Education with a major in social science. He reported that he had completed 10 hours in geography.

Summary of Contextual Variables

The four schools that participated in the study were similar in organization, administration, plant facilities, and student populations. All four schools and all 20 classes were racially integrated. Classrooms were self-contained. However, each teacher taught more than one class. Table 4.6 indicates the teacher and number of classes taught involved in this study.

Table 4.6

Teachers and the Number of Classes Taught

Teacher	Classes Taught
A	2
B	5
C	5
D	4
E	4
	Total = 20

The observed differences between the two treatment groups regarding the personal attributes of the teachers were deemed to be minor because all but one teacher taught classes in both treatments. Therefore, the researcher concluded that there were no contextual variables, other than treatment, that accounted for observed differences between the two treatments on the posttests.

Statistical Procedures

A 3 x 10 x 2, aptitude by classes-nested-within-treatments, by treatments, multivariate analysis of variance (MANOVA) was used with the learning, retention, and times-to-testing mean scores as the effects measures. This experimental design was used to determine if the differences between the mastery and non-mastery treatments produced differences ($p<.05$) in the average effects which were not the same at the high, middle, and low aptitude levels. The computer program used in the above analysis was the BMD 12V (Biomedical Computer Programs, 1973). This program can perform multivariate and univariate analyses of variance for any hierarchical design with cells that contained equal Ns, including nested, partially nested and partially crossed, and fully crossed designs. While a Multivariate, Univariate, Discriminate Analysis of Independent Data (MUDAID) program had been considered as the program for analysis of the data in this study it was found that it could not handle designs that included a nested factor. Consequently, the BMD 12V program was used because the independent variable, class, was nested within treatments.

Statement of the Statistical Hypotheses

The purpose of this study was to compare self-instructional mastery and non-mastery treatments to determine if there were differences in achievement and time of high, middle, and low aptitude students, using measures of learning, retention, and times-to-testing.

To accomplish this purpose, the following statistical hypotheses were tested at the .05 level of significance. The subscript order is the same as that used on the experimental layout Table 4.1, p. 69.

Hypotheses for MANOVA

1. Interaction: Treatment by Aptitude

$$H_0: \begin{bmatrix} \mu_{.111} \\ \mu_{.112} \\ \mu_{.113} \end{bmatrix} - \begin{bmatrix} \mu_{.121} \\ \mu_{.122} \\ \mu_{.123} \end{bmatrix} = \text{(Vectors of the high aptitude group by treatment)}$$

$$\begin{bmatrix} \mu_{.211} \\ \mu_{.212} \\ \mu_{.213} \end{bmatrix} - \begin{bmatrix} \mu_{.221} \\ \mu_{.222} \\ \mu_{.223} \end{bmatrix} = \quad \text{(Vectors of the middle aptitude group by treatment)}$$

$$\begin{bmatrix} \mu_{.311} \\ \mu_{.312} \\ \mu_{.313} \end{bmatrix} - \begin{bmatrix} \mu_{.321} \\ \mu_{.322} \\ \mu_{.323} \end{bmatrix} \quad \text{(Vectors of the low aptitude group by treatment)}$$

This null hypothesis states that the average difference between treatment effect vectors is the same at each aptitude level. This null hypothesis was tested against the two-tailed alternative hypothesis that the average difference between treatment effect vectors is not the same at each aptitude level.

2. Main Effects: Treatments

$$H_0: \begin{bmatrix} \mu_{..11} \\ \mu_{..12} \\ \mu_{..13} \end{bmatrix} = \begin{bmatrix} \mu_{..21} \\ \mu_{..22} \\ \mu_{..23} \end{bmatrix}$$

This null hypothesis states that with pupils pooled across the three levels of aptitude, there is no difference between the mastery and non-mastery treatment vectors of average effects. This null hypothesis was tested against the two-tailed alternative hypothesis that there is a difference.

3. Main Effects: Aptitude

$$H_0: \begin{bmatrix} \mu_{.1.1} \\ \mu_{.1.2} \\ \mu_{.1.3} \end{bmatrix} = \begin{bmatrix} \mu_{.2.1} \\ \mu_{.2.2} \\ \mu_{.2.3} \end{bmatrix} = \begin{bmatrix} \mu_{.3.1} \\ \mu_{.3.2} \\ \mu_{.3.3} \end{bmatrix}$$

This null hypothesis states that with pupils pooled across the two treatments, the vectors of achievement are the same at each of the three aptitude levels. This null hypothesis was tested against the two-tailed alternative hypothesis that with pupils pooled across the two treatments, the vectors of achievement are not the same at each of the three aptitude levels.

If statistical significance was found on the multivariate interaction hypothesis, an *a priori* decision was made to follow up by testing the univariate interaction hypotheses. No main effects were to be tested for, for the multivariate or univariate analyses. If there was no statistical significance on the multivariate interaction hypothesis, each of the multivariate main effects was to be tested. If these were statistically significant then the decision was to follow-up by testing each of the effects measures at the univariate level. If there were no statistically significant multivariate main effects then no follow-up tests were planned (Hummel and Sligo, 1971). Duncan's Multiple Range Test was the appropriate *post hoc* test for statistically significant outcomes for the univariate analyses (Edwards, 1968), while the Bonferroni t statistic was the appropriate *post hoc* test for simple effects (Marascuilo and Levin, 1970).

4. *Interaction: Treatment by Aptitude (Posttest)*

$$H_0: \mu_{.111} - \mu_{.121} = \mu_{.211} - \mu_{.221} = \mu_{.311} - \mu_{.321}$$

This null hypothesis states that the difference in average effects of the two treatments is the same at each aptitude level. This null hypothesis was tested against the two-tailed alternative that the difference in average effects of the two treatments is not the same at each aptitude level.

5. *Main Effects: Treatments (Posttest)*

H_o: $\mu_{..11} = \mu_{..21}$

This null hypothesis states that with pupils pooled across aptitude levels there are no statistical differences between treatments on the mean posttest scores. This null hypothesis was tested against the two-tailed alternative hypothesis that with pupils pooled across aptitude levels there are statistical differences between treatments on the mean posttest scores.

6. Main Effects: Aptitude (Posttest)

H_o: $\mu_{.1.1} = \mu_{.2.1} = \mu_{.3.1}$

This null hypothesis states that with pupils pooled across the two treatments there are no statistical differences between aptitude groups on the mean posttest scores. This null hypothesis was tested against the two-tailed alternative that with pupils pooled across the two treatments there are statistical differences between aptitude groups on the mean posttest scores.

The hypotheses for the analysis of variance for each of the effects measures followed the same format. Therefore, it was not necessary to state each set of hypotheses because of the repetition involved. The same hypotheses were applied to the measures of retention and times-to-testing.

Hypotheses for Simple Effects

13, 14, 15. Simple Effects: Learning

$$H_o: \begin{array}{l} \mu_{.111} = \mu_{.121} \\ \mu_{.211} = \mu_{.221} \\ \mu_{.311} = \mu_{.321} \end{array}$$

These null hypotheses state that at each of the three aptitude levels there is no difference between the two treatment means on the posttest. Each hypothesis was tested against its alternative that there are differences between treatment means across each level of aptitude on the posttest.

The hypotheses for simple effects for each of the effects measures followed the same format. Therefore, it was not necessary to state each set of hypotheses because of the repetition involved. The same hypotheses were applied to the measures of retention and times-to-testing.

Significance Level

In the present study the .05 significance level was used in testing the null hypotheses. This meant that a difference as large as or larger than the obtained one could occur by chance as infrequently as 5 times out of 100. Therefore, the probability of rejecting a true statistical hypothesis (Type 1 or α error) is .05.

A type II error (β) is the failure to reject a false statistical hypothesis. The relationship between α (Type 1 error) and β (Type II error) is inverse. Decreasing the probability of α Type 1 error increases the probability of a Type II error. The selection of a significance level, therefore, reflects a compromise between the relative importance of the two types of errors (Myers, 1966).

The power of a statistical test is defined as $1 - \beta$, or the probability of rejecting a statistical hypothesis when it is false and should be rejected. If α (Type 1 error) is held constant, the power of the significance test can be increased by increasing the number of observations in the sample; therefore, if power is increased, then the probability of β (Type II error) is decreased (Edwards, 1968).

By selecting a significance level of .05 instead of one that is higher (eg .01), the probability of making a Type II error is reduced. The .05 level, however, is strong enough to warrant concluding that the difference is not attributable merely to sampling errors. In cases where small sample sizes are used, Walker and Lev (1958) have stated that the level of significance should not be high because both factors reduce the power of a test. However, in the case of the present study where the sample size was larger (n=60), a .05 level of significance was considered appropriate.

Assumptions Underlying the Multivariate Analysis of Variance (MANOVA)

The MANOVA was used as one method of data analysis in this study because it was appropriate for testing the significance of differences between treatment means in terms of three dependent variables considered simultaneously (Tatsuoka, 1971). In order for MANOVA to be an appropriate test of the statistical hypotheses, the data must have met certain assumptions:

1. The variables under study must follow a multivariate normal distribution.

2. There must be equal dispersion matrices.

Assumptions Underlying the Analysis of Variance (ANOVA)

The Anova was used as the second method of data analysis in this study because it permitted a straight forward analysis of the hypothesis under consideration (Myers, 1966). In order for ANOVA to be an appropriate test of the statistical hypothesis, the data must meet the following assumptions:

1. The deviation, due to uncontrolled variability, of the individual mean scores from the treatment group population mean are independently distributed;

2. the deviation, due to uncontrolled variability, of the individual mean scores from the treatment group population mean are normally distributed;

3. the variance of the deviations, due to uncontrolled variability, is the same for all treatment group populations;

4. the null hypothesis is true (Myers, 1966).

If the first three assumptions are valid, then a significant F may be attributed to the falsity of the fourth assumption (Myers, 1966).

To meet the assumptions underlying the F test, the following procedures were used:

1. The validity of the independence assumption was met by the random assignment of classes to two groups and then random assignment of treatment to groups.

2. The validity of the normality assumption depended on the measure chosen by was of no concern since Norton (1953) had shown that the F ratio is little influenced by departures from normality (Myers, 1966).

3. The validity of the homogeneity of variance assumption was tested by using Hartley's test, and the data met this requirement for the achievement measure.

Limitations

The present study was limited to an investigation of the comparative effects of the self-instructional unit Functions of Cities in a mastery learning and non-mastery learning mode using three measures - learning, retention, and times-to-testing. It was further limited to the effects of materials which met the criteria specified by the researcher in Chapter 3.

A second limitation of the study was that while the researcher spent, on average, two days a week observing and assisting in the schools there could be no check made to ensure that written and oral directions were being carried out by the teachers and students in the study. Oral and written directions were provided prior to the beginning of treatment to each teacher along with a sample copy and classroom set of the materials in the format to be followed. Classrooms were visited regularly each week and teachers reported no irregularities. These procedures strengthened the assumption that the teachers and students followed the instructions outlined, but the degree to which individuals may have deviated from the established procedures cannot be determined.

A third limitation of the study was the use of an available pool of 539 seventh grade students in 20 classes in the Savannah-Chatham County School District. This population was not representative of a national sample. The subjects were below the national average in reading word knowledge as measured by the Iowa Tests of Basic Skills. In addition, the available pool of students did not follow the national ratio with regard to racial composition. In this study, approximately 48 per cent of the students were black. This percentage is considerably higher than the national percentage of 12.2 for metropolitan areas

(U. S. Bureau of Statistics, 1971).

A fourth limitation of the study was the necessity to use the same form of the measuring instrument to measure retention as was used to measure initial achievement. However, two aspects of the present study mitagated against a carry-over effect from the posttest. First, there was a planned 17 day interval between the administration of the posttest and the delayed posttest with no student feedback during this period. Second, students were not informed that they were to be retested.

A fifth limitation of the study concerned the lack of a pilot testing phase in the development of the materials and measuring instruments. This was due to the press of time in getting the materials into the schools for actual administration. However, a number of control steps were conducted that the researcher hoped would help offset the disadvantages of no pilot phase. The lack of a pilot test phase is acknowledged as another limitation.

A sixth limitation of the study was the use made of the mid-week discussion class. The activities that teachers and students indulged in were wide ranging. Originally, it was proposed that content-oriented activities would be used in the classroom. However, the researcher observed that slides, filmstrips, and films were utilized along with library activities. Many activities did not have specific bearing on the unit content. At other times students were permitted to work on their workbooks. This provided more time for students to work with the materials. Therefore, some students had more time to learn the content.

A seventh limitation of the study was the possibility that students did not respond independently. In order to maintain normal classroom learning conditions students were not reassigned to different desks in association with their respective aptitude levels. Consequently, there may have been interaction between students across aptitude groups in spite of the materials being self-instructional. An artificial, experimental environment was also avoided and the focus of the study, the effects upon achievement of students grouped by aptitude was able to be conducted and examined.

An eighth limitation was that the times-to-testing scores should have been transformed to eliminate the possibility that the ratios of the means and those of the variances were similar. The BMD 12V program was unable to transfer scores and consequently the mean cell scores and variances for times-to-testing may have been similar. This imposed a limitation for the data analysis.

Summary

This chapter presented a 3 x 10 x 2, aptitude by classes-nested-within-treatments, by treatments, multivariate analysis of variance (MANOVA) as the experimental design of the study. The main purpose of the study was to compare self-instructional mastery and non-mastery treatments to determine if there were differences in achievement and time of high, middle, and low aptitude students. Following the discussion of the experimental study, a description of the pattern of logic used in the study was provided. Factors that could not be controlled for statistically, were discussed and described as contextual variables.

Following the discussion of the contextual variables was a description of the procedures used in the experimental study and the limitations to the study. Data obtained in the experimental study were used to test the statistical hypotheses. The results of the tests within the limitations of the study are presented in the next chapter.

CHAPTER V

RESULTS AND DISCUSSION OF THE FINDINGS

The purpose of this chapter is to report, analyze, and discuss the data collected in the present study. The chapter is divided into two sections: 1) Presentation of the Findings; and 2) Discussion of the Findings. Tables 5.1, 5.2, and 5.3 present the raw cell mean data that was used in the multivariate analysis and subsequent data analyses.

Presentation of the Findings

The findings for the study are reported separately for each tested hypothesis.

Analysis of the Data by the Multivariate Analysis of Variance (MANOVA)

Analysis of the data by the BMD 12V program produced the following outcomes displayed in Tables 5.4, 5.5, 5.6, and 5.7.

Findings of Hypotheses for MANOVA

1. Interactions: Treatment by Aptitude

$$H_o: \begin{bmatrix}\mu_{.111}\\ \mu_{.112}\\ \mu_{.113}\end{bmatrix} - \begin{bmatrix}\mu_{.121}\\ \mu_{.122}\\ \mu_{.123}\end{bmatrix} =$$

(Vectors of the high aptitude group by treatment).

$$\begin{bmatrix}\mu_{.211}\\ \mu_{.212}\\ \mu_{.213}\end{bmatrix} - \begin{bmatrix}\mu_{.221}\\ \mu_{.222}\\ \mu_{.223}\end{bmatrix} =$$

(Vectors of the middle aptitude group by treatment).

Table 5.1

Learning: Cell Mean Scores by Treatment, Aptitude, and Class

	Class	1	2	3	4	5	6	7	8	9	10
MASTERY	High	$\overline{X}$ 41.00 S 9.4	20.00 8.5	43.80 6.7	39.00 16.0	49.32 3.9	46.60 11.7	35.50 16.5	39.78 12.6	39.33 17.2	28.43 8.7
	Middle	29.50 7.2	24.33 2.1	34.33 11.0	38.8 4.9	41.67 7.0	31.6 15.9	17.33 6.4	43.00 13.9	23.63 13.0	18.43 7.2
	Low	20.25 9.6	18.69 11.5	16.50 6.3	16.01 10.4	10.00 0.0	14.88 7.4	18.60 7.7	14.70 7.1	12.33 3.2	13.17 6.0
	Class	11	12	13	14	15	16	17	18	19	20
NON MASTERY	High	49.80 5.4	30.75 11.2	26.00 0.0	44.36 9.8	47.21 11.3	30.50 12.7	32.60 13.6	37.60 16.7	30.22 15.9	15.00 2.8
	Middle	36.33 10.6	24.40 3.4	17.63 7.6	31.60 12.1	27.33 13.4	16.13 9.0	16.60 4.2	24.80 9.7	28.00 4.6	15.50 6.3
	Low	19.17 8.0	17.25 6.5	12.46 3.4	31.50 30.4	21.00 0.0	20.50 7.4	15.11 7.3	16.44 8.2	10.13 5.6	12.00 5.4

Table 5.2

Retention: Cell Mean Scores by Treatment, Aptitude, and Class

	Class	1	2	3	4	5	6	7	8	9	10
MASTERY	High	$\bar{x}$ 43.80 s 10.4	32.50 3.6	44.40 8.7	39.00 12.0	49.00 6.4	48.60 3.3	33.67 14.2	42.67 13.5	43.67 7.4	31.67 13.3
	Middle	35.25 7.9	26.00 7.2	32.67 5.7	37.25 11.2	42.00 7.2	25.33 14.2	16.50 7.6	45.00 10.4	20.00 13.5	16.71 10.3
	Low	21.58 11.0	20.00 9.6	16.00 5.7	16.82 11.8	16.00 0.0	13.13 7.4	16.44 12.7	22.20 8.2	11.56 2.8	13.00 3.4
	Class	11	12	13	14	15	16	17	18	19	20
NON MASTERY	High	47.40 10.2	34.75 6.9	26.00 0.0	39.50 14.3	46.57 10.9	26.17 15.8	26.40 15.6	35.80 20.6	31.44 18.6	18.50 7.7
	Middle	36.33 10.4	17.00 9.8	18.38 7.5	21.60 9.1	26.33 11.9	12.13 2.6	13.40 4.9	23.00 10.4	28.00 10.5	12.00 8.5
	Low	17.50 6.5	15.13 10.5	12.91 5.7	26.00 25.5	8.00 0.0	16.50 4.1	13.33 4.5	17.11 10.5	11.25 10.0	11.90 9.5

Table 5.3

Times-to-Testing and Amount Completed: Cell Mean Scores by Treatment, Aptitude, and Class

	Class	1	2	3	4	5	6	7	8	9	10
MASTERY	High	8.60[1] $\bar{x}$ 611.80[2] s 84.5	6.50 487.50 53.0	7.00 500.18 86.5	5.67 520.00 82.7	8.86 561.27 106.1	8.00 541.20 108.8	7.67 625.33 187.6	7.11 646.11 159.0	8.33 530.00 54.8	8.29 352.71 211.1
	Middle	5.25 615.75 154.8	6.67 626.33 68.8	6.67 542.67 100.8	7.4 569.40 75.0	8.00 597.00 129.6	6.44 572.22 175.9	7.83 665.17 123.9	8.67 543.67 166.5	8.50 390.12 63.7	7.00 442.29 258.3
	Low	5.77 587.15 133.5	5.00 575.85 133.7	5.00 526.00 17.0	6.15 555.85 92.5	4.00 678.00 0.0	5.75 564.63 98.9	5.40 514.00 218.6	7.10 656.20 132.9	6.67 540.22 156.2	7.83 425.5 121.8
	Class	11	12	13	14	15	16	17	18	19	20
NON MASTERY	High	7.60 641.00 114.4	7.00 613.75 44.8	7.00 518.00 0.0	8.86 559.43 70.5	8.29 615.64 129.0	8.33 367.00 69.8	9.00 368.8 84.7	9.00 374.6 61.8	8.89 415.00 93.2	6.00 303.5 53.0
	Middle	7.67 662.67 41.3	8.20 481.40 107.8	5.88 452.00 50.7	7.67 593.40 85.9	8.33 687.50 78.3	7.50 376.50 166.5	8.2 279.8 120.5	7.80 495.6 140.7	8.67 388.65 115.9	9.00 358.00 56.6
	Low	5.67 637.83 79.0	7.00 516.25 83.0	6.72 414.64 76.8	7.5 572.00 185.3	9.00 567.00 0.0	8.40 361.8 137.8	8.89 496.56 170.5	8.40 439.00 104.9	7.88 372.88 81.4	8.10 363.4 105.4

1. Amount Completed 2. Times-to-Testing

Table 5.4

Multivariate Analysis of Variance Test of Significance - Learning, Retention, and Times-to-Testing

Source of Variance	Degree of Freedom	Approximate F Statistics	P<
Treatment	3, 34	14.8220	.001
Aptitude	6, 68	14.9901	.001
Class (Treatment)	54, 102	3.0536	
Treatment x Aptitude	6, 68	1.0163	NS*
Class x Aptitude (Treatment)			

*NS = Not significant p<.05

Table 5.5

Analysis of Variance for Treatment, Aptitude, and Interaction - Learning

Source of Variance	Sum of Squares	Degree of Freedom	Mean Square	F
Treatment	105.75	1	105.75	2.99
Aptitude	3987.03	2	1993.51	56.39*
Class (Treatment)	2060.33	18	114.46	3.24
Treatment x Aptitude	203.20	2	101.60	2.87
Class x Aptitude (Treatment)	1272.64	36	35.35	

*Indicates F ratios that are significant at the .05 level.

Table 5.6

Analysis of Variance for Treatment, Aptitude, and Interaction - Retention

Source of Variance	Sum of Squares	Degrees of Freedom	Mean Score	F
Treatment	553.65	1	553.65	16.28*
Aptitude	4540.26	2	2270.13	66.74*
Class (Treatment)	1782.13	18	99.01	2.91
Treatment x Aptitude	140.61	2	70.31	2.07
Class x Aptitude (Treatment)	1224.57	36	34.02	

*Indicates F ratios that are significant at the .05 level.

Table 5.7

Analysis of Variance for Treatment, Aptitude, and Interaction - Times-to-Testing

Source of Variance	Sum of Squares	Degrees of Freedom	Mean Square	F
Treatment	85744.19	1	85744.19	26.60*
Aptitude	1238.94	2	619.47	0.19
Class (Treatment)	455231.44	18	25290.63	7.85
Treatment x Aptitude	2162.13	2	1081.06	0.34
Class x Aptitude (Treatment)	116027.33	36	3222.98	

*Indicates F ratios that are significant at the .05 level.

$$\begin{bmatrix} \mu_{.311} \\ \mu_{.312} \\ \mu_{.313} \end{bmatrix} - \begin{bmatrix} \mu_{.321} \\ \mu_{.322} \\ \mu_{.323} \end{bmatrix}$$

(Vectors of the
low aptitude
group by treatment).

This statistical hypothesis, that the average difference between the two treatment vectors of average effects is the same at each aptitude level, was tested against the alternative hypothesis that the difference between the two treatment vectors of average effects is not the same at each aptitude level.

The multivariate F statistic for interaction of treatment and aptitude was not significant (see Table 5.4, p.104). The null hypothesis, therefore, was not rejected at the .05 level of significance. As the multivariate interaction null hypothesis was not rejected the univariate interaction hypotheses were not tested. Therefore, separate reports for the univariate analyses for the interaction hypotheses 4, 7, 10 are not presented.

2. Main Effects: Treatments

$$H_o: \begin{bmatrix} \mu_{..11} \\ \mu_{..12} \\ \mu_{..13} \end{bmatrix} = \begin{bmatrix} \mu_{..21} \\ \mu_{..22} \\ \mu_{..23} \end{bmatrix}$$

This statistical hypothesis, that with pupils pooled across the three levels of aptitude there is no significant differences between the mastery and non-mastery treatment vectors of average effects, was tested against the alternative hypothesis that there is a difference between the mastery and non-mastery treatment vectors of average effects.

The multivariate F statistic for the treatment effect was significant (see Table 5.4, p.104). Therefore the null hypothesis was rejected and the alternative hypothesis was accepted.

3. Main Effects: Aptitude

$$H_o: \begin{bmatrix} \mu_{.1.1} \\ \mu_{.1.2} \\ \mu_{.1.3} \end{bmatrix} = \begin{bmatrix} \mu_{.2.1} \\ \mu_{.2.2} \\ \mu_{.3.3} \end{bmatrix} = \begin{bmatrix} \mu_{.3.1} \\ \mu_{.3.2} \\ \mu_{.3.3} \end{bmatrix}$$

This statistical hypothesis, that across the two levels of treatments there are no differences between the three levels-of-aptitude vectors of average effects, was tested against the alternative hypothesis that across the two levels of treatments there are differences between the three levels-of-aptitude vectors of average effects.

The multivariate F statistic for aptitude effects was statistically significant (see Table 5.4, p.104). Therefore the null hypothesis was rejected and the alternative hypothesis was accepted.

Findings of Hypotheses for ANOVA

5. Main Effects: Treatments (Learning)

$$H_o: \mu_{..11} = \mu_{..21}$$

This statistical null hypothesis, that there is no statistical significant difference between treatments on the mean posttest scores, was tested against the alternative hypothesis that there are differences between treatments on the mean posttest scores.

The univariate F statistic for treatment was not significant (see Table 5.5, p.104). The null hypothesis, therefore was not rejected at the .05 level of significance.

6. Main Effects: Aptitude (Learning)

$$H_0: \mu_{.1.1} = \mu_{.2.1} = \mu_{.3.1}$$

This statistical null hypothesis, that there are no statistically significant differences among aptitude groups on the mean posttest scores, was tested against the alternative hypothesis that there are such differences.

The univariate F statistic for aptitude was significant (see Table 5.5, p.104).

To determine which pairs of aptitude means were significant the Duncan Miltiple Range Test was applied to the univariate cell matrix data to locate the source of the significant effect. Table 5.8 reports the results for aptitude effect on the posttest.

Table 5.8

Learning Mean Scores by Aptitude

Aptitude Group	N	Mean Score
High	20	36.34
Middle	20	27.05
Low	20	16.39

The results of the test are reported in Table 5.9. All differences were statistically significant.

Table 5.9

Learning: Summary of Results of the Duncan Multiple Range Test at the .05 Level of Significance for Aptitude Effect

Pairwise Comparisons	Significance
1 VS 2	.05
1 VS 3	.05
2 VS 3	.05

8. Main Effects: Treatment (Retention)

H_0: $\mu_{..12} = \mu_{..22}$

This null hypothesis, that there is no difference statistically between treatments on the mean delayed posttest scores, was tested against the alternative hypothesis that there is such a difference.

The univariate F statistic for treatment was significant (see Table 5.6, p.105). The univariate cell matrix was examined to determine the cell of significant treatment. Treatment 1 (mastery) was significantly larger than treatment 2 (non-mastery). Table 5.10 shows the difference between the two treatment means.

Table 5.10

Retention: Mean Scores for Treatments

	Mastery Treatment 1	Non-Mastery Treatment 2
n	30	30
Means	29.05*	22.98

*The difference between the mean is significant at the .05 level of significance.

9. Main Effects: Aptitude (Retention)

H_0: $\mu_{.1.2} = \mu_{.2.2} = \mu_{.3.2}$

This null hypothesis, that there are no statistically significant differences among aptitude groups on the mean delayed posttest scores, was tested against the alternative hypothesis that there are such differences.

The univariate F statistic for aptitudes was significant (see Table 5.6, p.105).

To determine which pair of aptitude means was significant the Duncan Multiple Range Test analysis was applied to the univariate cell matrix data to determine the source of the significant effect. Table 5.11 reports the results for aptitude effect on the delayed posttest.

Table 5.11

Retention: Mean Scores by Aptitude

Aptitude Groups	Cell Size	Mean Score
1	20	37.03
2	20	25.24
3	20	15.77

The results of the test are reported in Table 5.12. All differences were statistically significant.

Table 5.12

Retention: Summary of Results of the Duncan
Multiple Range Test at the .05 Level of Significance
for Aptitude Effect

Pairwise Comparisons	Significance
1 VS 2	.05
1 VS 3	.05
2 VS 3	.05

11. Main Effects: Treatment (Times-to-Testing)

H_0: $\mu_{..13} = \mu_{..23}$

This null hypothesis, that there is no statistically significant difference between treatments on mean times-to-testing, was tested against the alternative hypothesis that there is such a difference.

The univariate F statistic for treatment was significant (see Table 5.7, p.105). The univariate cell matrix was examined to determine the cell of significant treatment. Treatment 2 (non-mastery) took significantly less times-to-testing than Treatment 1 (mastery). Table 5.13 shows the difference between the two treatment means.

Table 5.13

Times-to-Testing in Minutes: Mean Number of Minutes for Treatments

	Mastery Treatment 1	Non-Mastery Treatment 2
n	30	30
Means	551.67	476.06*

*The difference between these means is significant at the .05 level of significance.

12. Main Effects: Aptitude (Times-to-Testing)

H_0: $\mu_{.1.3} = \mu_{.2.3} = \mu_{.3.3}$

This null hypothesis, that there are no statistically significant differences among aptitude groups on the mean times-to-testing scores, was tested against the alternative hypothesis that there are such differences.

The univariate F statistic for aptitude was not statistically significant (see Table 5.7, p. 105). The null hypothesis, therefore, was not rejected at the .05 level of significance.

Findings of Hypotheses for Simple Effects

13, 14, 15. Simple Effects: Learning

H_0: $\mu_{.111} = \mu_{.121}$

$\mu_{.211} = \mu_{.221}$

$\mu_{.311} = \mu_{.321}$

These null hypotheses state with respect to learning that at each aptitude level there is no significant difference between the mastery and non-mastery treatment means.

To test for significance between treatments across each level of aptitude the appropriate post hoc technique was the Bonferroni t test. Marascuilo and Levin (1970) suggest that manipulations of the sources of variation and degrees of freedom in what is called a nested or simple effects design and tested with the appropriate post hoc technique will provide the necessary information. They claim that,

> "From this, one may justly infer that sums of squares and degrees of freedom, like matter, are neither created nor destroyed, but are merely revealed in different forms."

To test the simple effects hypothesis a conservative alpha (α) of .10 was selected and partitioned into three equal sub-parts for each of the hypothesis. Therefore .10/3 or .033 was the significance level used.

As each set of hypotheses used the same components, they can be stated here as:

EW = the alpha (α) level = .10

df error = 36

n = 10

Bonferroni t = 2.215

The formula for computing the contrast is

$$S_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{2 \text{ M.S. error}}{N}}$$

where,

$M.S._{error}$ = mean square for error.

Bonferroni t statistic = $\sqrt{\frac{2\ M.S._{error}}{n}} \cdot$ Bonferroni t.

The Bonferroni t test tests for differences between the cell means. The cell means for treatments across each level of aptitude are presented in Table 5.14.

Table 5.14

Learning: Cell Means and Differences for Treatments Across Each Level of Aptitude

	Mastery Treatment 1	Non-Mastery Treatment 2	Differences
High	38.28	34.40	3.87
Middle	30.26	23.83	6.43*
Low	15.22	17.56	-2.34

*Significant at the .05 level of significance.

Application of the Bonferroni t test yielded a critical value of 5.89. Therefore, a difference as large as 5.89 was significant. The results of testing the hypotheses follow.

The null hypothesis that, with respect to learning, there is no significant difference between the mastery and non-mastery treatment means for high aptitude students was not rejected. High aptitude mastery treatment students did not differ significantly from high aptitude non-mastery treatment students on the posttest measure.

The null hypotheses that, with respect to learning, there is no significant difference between the mastery and non-mastery treatment means for middle aptitude students was rejected in favour of the alternative hypothesis. The posttest treatment mean for middle aptitude mastery students was significantly higher than the posttest treatment mean for middle aptitude non-mastery students.

The null hypothesis that, with respect to learning, there is no significant difference between the mastery and non-mastery treatments means for low aptitude students was not rejected. Low aptitude mastery students did not differ significantly from low aptitude non-mastery treatment students on the posttest measure.

16, 17, 18. Simple Effects: Retention

$$H_0: \begin{array}{ccc} \mu_{.112} & = & \mu_{.122} \\ \mu_{.212} & = & \mu_{.222} \\ \mu_{.312} & = & \mu_{.322} \end{array}$$

These null hypotheses state with respect to retention, that at each aptitude level there is no significant difference between the mastery and non-mastery treatment means.

The Bonferroni t test of significance was used to test for significance between treatments across each level of aptitude. See the description of the Bonferroni t test on p. 113. The cell means for treatments across each level of aptitude are presented in Table 5.15.

Table 5.15

Retention: Cell Means and Differences for Treatments Across Each Level of Aptitude

	Mastery Treatment 1	Non-Mastery Treatment 2	Differences
High	40.82	33.25	9.56*
Middle	29.67	20.82	8.85*
Low	16.67	14.86	1.81

*Significant at the .05 level of significance.

Application of the Bonferroni t test yielded a critical value of 5.78. Therefore, a difference as large as 5.78 was significant. The results of testing the hypotheses follow.

The null hypothesis that, with respect to retention, there is no significant difference between the mastery and non-mastery treatment means for high aptitude students was rejected in favour of the alternative hypothesis. The delayed posttest treatment mean for high aptitude mastery students was significantly higher than the delayed posttest treatment mean for high aptitude non-mastery students.

The null hypothesis that, with respect to retention, there is no significant difference between the mastery and non-mastery treatment means for middle aptitude students was rejected in favour of the alternative hypothesis. The delayed posttest treatment mean for high aptitude mastery students was significantly higher than the delayed posttest treatment mean for middle aptitude non-mastery students.

The null hypothesis that, with respect to retention, there is no significant difference between the mastery and non-mastery treatment means for low aptitude students was not rejected. Low aptitude mastery treatment students did not differ significantly from low aptitude non-mastery treatment students on the delayed posttest measure.

19, 20, 21. Simple Effects: Times-to-Testing

$$H_0: \begin{array}{ccc} \mu_{.113} & = & \mu_{.123} \\ \mu_{.213} & = & \mu_{.223} \\ \mu_{.313} & = & \mu_{.323} \end{array}$$

These null hypothesis state, with respect to times-to-testing, that at each aptitude level there is no significant difference between the mastery and non-mastery treatment means.

The Bonferroni t test of significance was used to test for significance between treatments across each level of aptitude. See the description of the Bonferroni t test on p. 113. The cell means for treatments across each level of aptitude are presented in Table 5.16.

Table 5.16

Times-to-Testing: Cell Means in Minutes and Differences for Treatments Across Each Level of Aptitude

	Mastery Treatment 1	Mastery Treatment 2	Differences
High	537.20	477.67	59.52*
Middle	556.46	477.55	78.91*
Low	561.33	472.96	88.38*

*Significant at the .05 level of significance.

Application of the Bonferroni t test yielded a critical value of 17.80. Therefore, a difference as large as 17.80 was significant. The results of testing the hypotheses follow.

The null hypothesis that, with respect to times-to-testing, there is no significant difference between the mastery and non-mastery treatment means for high aptitude students was rejected in favour of the alternative hypothesis. The times-to-testing mean score for high aptitude mastery students was significantly greater than the times-to-testing mean score for high aptitude non-mastery students.

The null hypothesis that, with respect to times-to-testing, there is no significant difference between the mastery and non-mastery treatment means for middle aptitude students, was rejected in favour of the alternative hypothesis. The times-to-testing mean score for middle aptitude mastery students was significantly greater than the times-to-testing mean score for middle aptitude non-mastery students.

The null hypothesis that, with respect to times-to-testing, there is no significant difference between the mastery and non-mastery treatment means for low aptitude students was rejected in favour of the alternative hypotheses. The times-to-testing mean score for the low aptitude mastery students was significantly greater than the times-to-testing mean score for the low aptitude non-mastery students.

Discussion of the Findings

This study found that differences between aptitude levels were increased rather than diminished when self-instructional materials were used. High aptitude students learned and retained more of the geography unit than middle or low aptitude students, while middle aptitude stu-

Table 5.17

Summary of Multivariate and Univariate Tests of Significance: Interaction and Main Effects

Statistical Hypotheses (Null)	F	Level of Significance
There are no differences:		
I. Between vectors (MANOVA) of learning, retention, and times-to-testing;		
1. Interaction: treatment by aptitude	1.02	N.S.
2. Main Effects: treatment	14.82	.001
3. Main Effects: aptitude	14.99	.001
II. Learning (ANOVA): mean differences for interaction and main effects;		
4. Interaction: treatment by aptitude	2.87	N.S.
5. Main Effects: treatment	2.99	N.S.
6. Main Effects: aptitude	56.39	N.S.
III. Retention (ANOVA): mean difference for interaction and main effects;		
7. Interaction: treatment by aptitude	2.07	N.S.
8. Main Effects: treatment	16.28	.05
9. Main Effects: aptitude	66.74	.05
IV. Times-to-Testing (ANOVA): mean differences for interaction and main effects;		
10. Interaction: treatment by Aptitude	0.34	N.S.
11. Main Effects: treatment	26.60	.05
12. Main Effects: aptitude	0.19	N.S.

Table 5.18

Summary of Tests of Significance for Simple Effects: Comparisons of Aptitude Levels Across Treatments

Statistical (Null) Hypotheses	Mean Score Mastery Treatment	Mean Score Non-Mastery Treatment	Mean Difference	Level of Significance
There are no differences:				
I. Learning: treatment means across aptitude levels.				
(13) High	38.28	34.40	3.87	N.S.
(14) Middle	30.26	23.83	6.43	.05
(15) Low	15.22	17.56	-2.34	N.S.
(Simple Effects of II Table 5.17)				
II. Retention: treatment means across aptitude levels.				
(16) High	40.82	33.25	7.56	.05
(17) Middle	29.67	20.82	8.85	.05
(18) Low	16.67	14.86	1.81	.05
(Simple Effects of III Table 5.17)				
III. Times-to-Testing: treatment means across aptitude levels.				
(19) High	537.20*	477.67*	59.53*	.05
(20) Middle	556.46*	477.55*	78.91*	.05
(21) Low	561.33*	472.96*	88.38*	.05
(Simple Effects of IV Table 5.17)				

*Expressed in minutes

dents learned and retained more of the geography unit than low aptitude students. These results suggest that achievement was a function of the capacities and talents for learning that students of varying aptitude brought to the instruction.

The analysis of simple effects of treatments across each level of aptitude found that the mastery treatment facilitated greater retention for the high and middle aptitude students, and greater learning for the middle aptitude students. This was accomplished due to the feedback correction procedures required of the mastery students and the increased time that these procedures required of the mastery students for relearning. This result is consistent with that of Fishburne (1971) who used a programmed and non-programmed text. He found that exposure to the programmed text increased learning and retention but took more time across levels of reading. He attributed increased student learning to the extra time taken with the materials. Therefore, it would appear that self-instructional materials at least facilitate retention for students of high and middle aptitude students. However, the mastery procedures did not facilitate learning and retention for low aptitude students.

Low aptitude mastery students neither learned nor retained the geography material more than low aptitude non-mastery students. The low aptitude students used in this study obtained very low reading scores as measured by the Iowa Tests of Basic Skills. When converted to grade equivalent scores the low aptitude mastery and non-mastery students were reading at approximately fourth grade level. This is almost four grade levels below actual classroom level (see Tables 4.3, 4.4, and 4.5 in Chapter Four) and at least two grades below the Grade 6 reading level of the materials Functions of Cities used in the study (see p.55 and Table 3.2 for discussion of readability). Therefore, the lack of differences between the low aptitude mastery and non-mastery students can be explained by the lack of verbal facility that low aptitude students brought to instruction. This was particularly evident in the scores obtained on the 40 item multiple choice and the 24 item recall chart tables appearing in Chapter 3. The low aptitude mastery and non-mastery students consistently scored lower than the middle and high aptitude groups on the 40 item multiple choice test and often did not start the 24 item recall test (see Tables 3.4 and 3.5). This strongly suggests that the strength of learning by low aptitude students was indeed low. Another factor that reinforces this position is that there was only a one chapter difference between high and low aptitude students at the completion of instruction. This suggests that low aptitude students did not spend the necessary time in relearning the material necessary to improve their learning. The difficulty of the material due to their inherent reading and vocabulary deficiencies probably caused frustration in learning and reduced their task orientation. Therefore, the materials Functions of Cities were probably too difficult for low aptitude students.

The review of the nine studies comparing mastery to non-mastery strategies revealed that two were below the college level, three used self-instructional materials, and none used social science materials. Within this context, all studies reported that mastery facilitated learning more than a non-mastery treatment. The emphasis of research was at the university or college level where students used could not be considered a representative sample of normal classroom conditions.

The results of the present study indicate that when self-instructional mastery procedures are used they do not facilitate greater post-test average performance than non-mastery procedures. The findings are contrary to Moore, Mahan and Ritts (1968), Green (1969), and Gentile (1970). These researchers used self-paced procedures. However, they used content that is sequential by nature (math and science content) and each learning task was contiguous with the next. This study used geography materials organized in a specific sequence devised by the researcher. However, the materials were constructed and organized around two major generalizations and this scheme was followed through each of the chapters. The results of the present study apply to the materials and students in this study but it is reasonable to suppose that similar results would be obtained if the same materials were used with students who contained similar contextual characteristics.

The literature concerning retention (Block, 1970; Kersh, 1970; Romberg, Shepler, and King, 1970; and Wentling, 1970) found that retention is facilitated when group-paced instruction is used with correction and feedback. This study found that when self-instructional geography materials were used mastery procedures facilitated greater retention than non-mastery procedures as measured by the delayed post-test. Therefore, this would suggest that the correction-feedback procedures, either group-paced or self-instruction; facilitated greater retention of original learning.

The literature review showed that only two studies reported the time variable (Merrill, Barton, and Wood, 1970; Block, 1970). Both studies indicated that learning became increasingly efficient over a series of sequenced learning units in class-paced instruction. This study did not support these findings. Mastery students used considerably more time to learn the material than non-mastery students. These time differentials also increased when comparisons were made between aptitude levels. Therefore, the results of this study would suggest that self-paced mastery instruction requires more time than self-paced non-mastery instruction or class-paced instruction.

This chapter has presented the findings of the study for each of the statistical hypotheses and has discussed some of the implications. The next chapter provides a summary of the study, introduces some educational implications, and recommends areas for further research.

CHAPTER VI

SUMMARY, DISCUSSIONS, AND RECOMMENDATIONS

Summary

This study was conducted under the sponsorship of the Geography Curriculum Project of the University of Georgia. The purpose of the study was to determine the effects of a self-instructional mastery procedure upon the average achievement of students of varying aptitudes using measures of learning, retention, and times-to-testing.

Research Hypotheses

The major purpose of this study was to compare self-instructional mastery and non-mastery treatments to determine if there were differences in learning, retention and time-to-testing of high, middle and low aptitude students.

The following research hypotheses were investigated.

1. The mastery and non-mastery treatments will produce differences in the average affects which are not the same ($p<.05$) at the high, middle, and low aptitude levels measured by posttests of:

(a) learning

(b) retention

and a measure of,

(c) times-to-testing

2. With pupils pooled across the three levels of aptitude the difference between the mastery and the non-mastery treatments will pro-

duce differences ($p<.05$) in the average achievement measured by geography posttests of:

(a) learning

(b) retention

and a measure of,

(c) times-to-testing

3. With pupils pooled across the two treatments, there are differences among the three levels-of-aptitude vectors of average effects ($p<.05$) measured by geography posttests of

(a) learning

(b) retention

and a measure of

(c) times-to-testing

Procedures

A geography unit titled Functions of Cities was developed by the researcher. The self-instructional unit consisted of a student text and two forms of the student workbook. Two treatments were devised. The non-mastery treatment (T_2) received the student text and a workbook. The workbook contained prescribed activities and a single review test for each chapter. Students worked through both. The mastery treatment (T_1) received the same student text but the workbook varied. Each chapter of the workbook contained two review tests. If the criterion level was not attained in the first review test, mastery students were required to correct and relearn material and then take a second review test.

Two basic concepts of urban geography used in relations to cities, function and economic base, were identified as the major themes of these project materials. The two major concepts along with related generalizations and facts were recorded in a table of specifications which was used in the construction of the measuring instruments.

A 40 item multiple choice test and a 24 item recall test was developed by the researcher to collect data to measure students' performance for the experiment. Both tests were used to measure learning and retention of the content materials. The retention measure was administered 17 days after the conclusion of instruction.

Twenty grade seven classes from the Savannah-Chatham County School District served as the experimental population. Treatments were randomly assigned to classes in each school. All subjects were administered the word meaning section of the _Iowa Tests of Basic Skills: Form 5 and 6_ (Lindquist and Hieronymus, 1971). Students within the 20 classes were then placed in three levels of aptitude. Classes were then randomly assigned to two groups and treatment was randomly assigned to groups.

Because individual classes were the smallest units of independence, class should have been the smallest unit of analysis. However, because this study focused upon aptitude groups within class, the aptitude group mean was used as the analysis unit. The mean was obtained from the unequal Ns for each of the sixty cells. A 3 x 10 x 2,aptitude by classes-nested-within-treatments, by treatments, multivariate analysis of variance was used to compare the differential effects of two treatments across three levels of aptitude.

Findings

The findings of the investigation were reported separately for each of the statistical hypotheses used to test the research hypotheses. The research hypotheses were intended to establish whether self-instructional, mastery procedures reduced differences in achievement of high, middle, and low aptitude students, as measured by tests of learning, retention, and times-to-testing.

✓ This study found that differences in aptitude were not reduced when self-instructional materials were used. The findings are reported, more specifically, for interaction of treatment and aptitude, in terms of the main effects (treatment and aptitude), and simple effects of aptitude levels across treatments for learning, retention, and times-to-testing.

Findings of the Treatment by Aptitude Interaction

No significant interactions between treatment and aptitude levels were found on the learning, retention, and times-to-testing measures. Treatment and aptitude were not acting together in this study.

Findings Between Treatment Groups

Students of high aptitude scored significantly higher than middle and low aptitude students as did students of middle aptitude over students of low aptitude on learning and retention. However, there were no differences on the times-to-testing between any of the aptitude levels.

Findings of the Aptitude Levels Across Treatment: Simple Effects

High and middle aptitude mastery treatment students retained more than high and middle aptitude non-mastery treatment students and middle aptitude mastery students learned more than middle aptitude non-mastery students. There was no difference between learning and retention for the low aptitude students across treatments. However, high, middle, and low aptitude non-mastery students used less time than high, middle, and low aptitude mastery treatment students.

Discussion of Educational Implications

The basic concerns of the researcher in this study were the effects that a self-instructional mastery procedure had on students of varying aptitude when social science materials were used. Since the study found that the mastery procedrue did not facilitate learning and retention for low aptitude students the following suggestions would seem in order.

The disadvantaged learner brings to the classroom many learning problems. It should be the teacher's and the school's responsibility to assist these students. Mastery procedures would appear to offer the disadvantaged student some hope of overcoming some of their environmental and hereditary learning deficiencies if a teacher is prepared to work closely with the student and to carefully monitor the mastery procedure at each level. The lack of teacher monitoring in administering the review tests may have contributed to the poor performance of the low aptitude students. The second review test for the mastery students can be a strong relearning tool if used correctly. The researcher did not request that the teachers monitor the retaking of the review test. The researcher believes that this led to only cursory examination of the learning material by all students and particularly low aptitude students. This is a weakness in the procedures used in this study and the researcher strongly recommends that this be controlled for in subsequent studies of a similar design and nature to this one. While the results of this study do not support the use of self-instructional mastery materials with the low aptitude student, class-paced mastery materials may operate more successfully with the slow learner.

The lack of success by low aptitude students was also a function of the degree to which low aptitude students were task oriented. Typically, low aptitude students require close personal supervision by the teacher, frequent feedback, and learning success. Stuempfig and Maehr (1970) found in a study concerning matching of materials and student characteristics, that low performing students performed better with personal rather than impersonal feedback. The low aptitude students, in this study, used self-instructional materials where all students responded independently to the learning exercises. As the low aptitude students performance, as measured by the geography achievement test, did not differ from chance to any great degree, this strongly suggests that self-instructional materials do not operate as well with low aptitude students as they operated with middle and high aptitude students.

The purpose of including the time measure in the study was to determine whether the use of correction feedback procedures which required more time facilitated learning across levels of aptitude. As the correction-feedback procedures required that more time be spent by the mastery students it was expected that mastery students should have increased achievement. However, there were two disadvantages to this practice. First, the mastery students did not complete as much of the unit as non-mastery students. Therefore, the advantage of superior achievement must be weighed against the disadvantage of less work covered. The school must decide where its priority lies in this regard. Second, the learning of social science materials and other disciplines compete for a students learning time each day of his school life. In a society where success is most often measured by quantity rather than quality, schools may not be able to afford the extra time that a mastery procedure appears to require. The economics of achievement as weighed against extra time to attain quality of learning may not be compatible in today's schools.

Recommendations for Further Research

Based on the findings and conclusions of the present study, the researcher submits the following specific recommendations for further systematic research relating to the effects of mastery on students of varying aptitude.

The first recommendation for further research can be found in the threats to external validity which were inherent in the procedures and design of this study. The reactive arrangement of treatment was a possible limitation of the present study. Therefore, the following recommendations are made for further research:

1. This study should be replicated in its present form using a larger number of schools, grade levels, and school systems.

2. This study should be replicated in its present form without the artificiality of an experimental setting and without the student's knowledge that he is involved in an experiment.

The first recommendation made above emphasizes the fact that the sample used in the experiment was drawn from a population of seventh grade students of the Savannah-Chatham County School System. Thus, the findings of this study can only be generalized to similar populations that have similar characteristics. The second recommendation emphasizes the fact that the sample used in the present study may have realized the experimental nature of their situation. Future research should control for this reactive arrangement.

The third recommendation concerns the select of material and its implementation in the classroom. The materials Functions of Cities should be used in a subsequent study where the administration of the materials are closely monitored by the researcher. This would overcome problems that Gaines (1971), Kim (1969) and this study encountered in making the mastery treatment more potent. Ideally, the researcher should live on site for the period of the study. Therefore, the following recommendations are made for further research:

3. The unit Functions of Cities should be administered with greater researcher control and supervision to ensure that the differences between treatments is enhanced.

4. A study should be conducted where the review tests and answer sheets are distributed by the teacher when the student has demonstrated that he is ready to perform these tasks.

This recommendations would assist the teacher and the researcher to more closely monitor the treatment in the classroom. It was suspected in this study that the availability of the review tests and the answer sheets may have contributed to the formation of slovenly learning habits.

The fifth recommendation suggests that the unit Function of Cities be used with various class levels and that a class-paced procedure be devised to observe the effects of treatment across aptitude levels. Therefore, the following recommendation is made for further research:

5. A class-paced procedure should be devised for various class levels to observe the effects of treatment across varying aptitude levels.

This recommendation was made because of the disadvantage that low aptitude students confronted in this study. Closer personal student-teacher contact may assist the low aptitude students to overcome some of their personal weaknesses such as poor vocabulary, poor understanding of the content, and frustration with the procedures.

Summary of Recommendations

The need for further research comparing self-instructional, mastery procedures with self-instructional non-mastery procedures in student's performance as measured by tests of learning, retention, and times-to-testing has been demonstrated.

The findings of the present study are generalizable only to similar populations using similar instructional materials and measuring learning outcomes using similar measuring instruments to those used in the study.

The suggestions for further research recommended previously are beyond the capabilities of any single researcher working alone to accomplish. A systematic comprehensive study of mastery is needed. This would entail a large scale, well coordinated team effort where individual investigators would each focus on a single task or variable yet coordinate his research with that of his colleagues. A trend beginning with this study, has begun at the University of Georgia where a series of studies have been planned. It is strongly believed by this researcher that such a group effort is needed, not only for research in the broad spectrum of mastery, but in the many aspects of investigating theories and practices in education.

REFERENCES

Airasian, P. W. An application of a modified version of John B. Carroll's model of school learning. Unpublished Master's thesis, University of Chicago, 1967.

Ammons, R. B. and Ammons, H. S. Full picture vocabulary test. Missoula, Montana: Psychological Test Specialist, 1948

Ausubel, D. P. The psychology of meaningful verbal learning. New York: Grune & Stratton, 1963.

Biehler, R. F. A first attempt at a 'learning for mastery' approach. Educational Psychologist, 7, No. 3 (1970), 7-9.

Block, J. H. The effects of various levels of performance on selected cognitive, affective and time variables. Unpublished Ph.D. dissertation, University of Chicago, 1970.

Block, J. H. Mastery learning: Theory and practice. New York: Holt, Rinehart and Winston, Inc., 1971.

Block, J. H. Mastery learning in the classroom: An overview of recent research. Paper presented at the American Educational Research Association, New Orleans, 1973.

Bloom, B. S. (Ed.), Englehart, M. D., Furst, E. J., Hill, W. H., and Krathwohl, D. R. Taxonomy of educational objectives: The classification of educational goals. Handbook 1: Cognitive domain. New York: David McKay Company, 1956.

Bloom, B. S. Learning for mastery, Evaluation Comment, 1, No. 2 (1968).

Bloom, B. S., Hastings, T. J., and Madaus, G. F. Handbook on formative and summative evaluation of student learning. New York: McGraw-Hill Book Company, 1971.

Boten, H. A teaching-testing scheme for senior high school American history. Classroom and administrative problems in the teaching of the social sciences. Second Yearbook of the National Council of the Social Studies. Philadelphia: McKinley, 1932.

Boyington, G. Experiments with diagnostic tests to determine knowledge of study tools and techniques in the social studies. Classroom and administrative problems in the teaching of the social sciences. Second Yearbook of the National Council of the Social Studies. Philadelphia: McKinley, 1932.

Brownell, W. A. Criteria of learning in educational research. Journal of Educational Psychology, 39 (1948), 170-182.

Campbell, D. T., & Stanley, J. C. Experimental and quasi-experimental designs for research. Chicago: Rand McNally (Reprint from Handbook of research on teaching, Gage, N. L.), 1963.

Carroll, J. B. A model of school learning, Teacher's College Record, 64 (1963), 723-33.

Carroll, J. B., and Spearritt, D. A study of a model of school learning. Monograph No. 4. Cambridge, Massachusetts: Harvard University, Center for Research and Development of Educational Differences, 1967.

Collins, K. M. A strategy for mastery learning in freshman mathematics. Unpublished study, Purdue University, Division of Mathematical Sciences, 1969.

Dale, J. R. An analysis of the effects on achievement using the forced inferential response mode in an intermediate grade population geography unit. Unpublished Ed.D dissertation, University of Georgia, 1972.

Davis, F. B. Educational measurements and their interpretation. Belmont, California: Wadsworth Publishing Company, 1964.

DeCecco, J. P. The psychology of learning and instruction: Educational psychology. Englewood Cliffs, New Jersey: Prentice Hall, 1968.

Dixon, W. J. (Ed.), Biomedical Computer Programs. Los Angeles: University of California Press, 1973.

Dumbleton, D. D. The effects of guided inquiry and expository materials on cognitive learning, retention, and transfer in a social studies unit for secondary level students. Unpublished Ed.D. dissertation, University of Georgia, 1973.

Dunn, L. M. Peabody picture vocabulary test. Minnesota: American Guidance Services, Inc., 1959.

Edwards, A. L. Experimental design in psychological research. New York: Holt, Rinehart & Winston, 1968.

Fishburne, R. P., Jr. The comparison of a programmed and a non-programmed text on evolution for the fifth grade. Unpublished Master's thesis, Georgia Southern College, 1971.

Flesch, R. How to test readability. New York: Harper, 1949.

Frech, W. P., Jr. An analysis of the Anthropology Curriculum Project materials *The Concept of Culture*, on the ethnocentric attitudes of fourth grade students. Unpublished Ed.D. dissertation, University of Georgia, 1973.

Gaines, W. G. An application of John B. Carroll's model of school learning to the teaching of anthropology. Unpublished doctoral dissertation, University of Georgia, 1971.

Gentile, J. R. A mastery strategy for introductory educational psychology. Unpublished materials, State University of New York at Buffalo, Department of Educational Psychology, 1970.

Glaser, R. Adapting the elementary curriculum to individual performance. *Proceedings of the 1967 Invitational Conference on Testing Problems*. Princeton: Educational Testing Service, 1968, 3-36.

Glass, G. V., & Stanley, J. C. *Statistical methods in education and psychology*. New Jersey: Prentice-Hall, 1970.

Green, B. A. *A self-paced course in freshman physics*. Cambridge, Massachusetts: Massachusetts Institute of Technology, Education Research Center, 1969.

Greene, W. W., Jr. Anthropological teaching in the first and fourth grades: A comparison of trained and non-trained teachers as measured by pupil test performance. Unpublished Ed.D. dissertation, University of Georgia, 1966.

Gronlund, N. E. *Preparing criterion - referenced tests for classroom instruction*. New York: The Macmillan Co., 1973.

Hieronymus, A. N. and Lindquist, E. F. Manual for administrators, supervisors, and counselors. Boston: Houghton Mifflin Co., 1974.

Hummell, T. J., & Sligo, J. R. Empirical comparisons of univariate and multivariate analysis of variance procedures. *Psychological Bulletin*, 1971, *76*, 49-57.

Jones, F. G. Functions of cities. Geography Curriculum Project, Publicaiton No. 74-1, University of Georgia.

Keller, F. S. "Goodbye, Teacher ...," *Journal of Applied Behavioral Analysis*, *1* (1968), 79-89.

Kersh, M. E. A strategy for mastery learning in fifth-grade arithmetic. Unpublished Ph. D. dissertation, University of Chicago, 1970.

Khan, S. B. "Affective correlates of academic achievement." *Journal of Educational Psychology*, 60, (1969), 216-21.

Kim. H. Learning rates, aptitudes, and achievements. Unpublished Ph.D. dissertation, University of Chicago, 1970.

Kim H. *A study of the Bloom strategies for mastery learning*. Seoul: Korean Institute for Research in the Behavioral Sciences, 1970.

Kim, H. *The mastery learning project in the middle schools*. Seoul: Korean Institute for Research in the Behavioral Sciences, 1970.

Kim, H. *Mastery learning in the Korean middle schools*, U.N.E.S.C.O. Regional Office for Education in Asia, 6, No. 1 (September 1971), Sec. 1, 55-60.

Lee, Y. D. *Interaction improvement studies on the mastery learning project - final report on mastery learning program, April - November, 1971*. Research Center, Seoul National University November, 1971.

Lindquist, E. F., & Hieronymus, A. N. *Iowa tests of basic skills*, Forms 5 and 6. Boston, Mass.: Houghton Mifflin Co., 1971.

Maracuilo, L. A., & Levin, J. R. Appropriate post hoc comparisons for interaction and nested hypotheses in analysis of variance designs: The elimination of type IV errors. *American Educational Research Journal*, 1970, 7 (3), 397-421.

Mayo, S. T., Hunt, R. C., & Tremmel, F. A mastery approach to the evaluation of learning statistics. Paper presented at the annual meeting of National Council on Measurement in Education, Chicago, Illinois, 1968.

Merrill, M. D., Barton, K., & Wood, L. E. Specific review in learning a hierarchical imaginary science, *Journal of Educational Psychology*, 61, 1970, 102-9.

Mitchell, J. J. A critical analysis of the research on mastery learning, Unpublished Master's thesis, University of Georgia, 1974.

Moore, J. W., Mahan, J. M., & Ritts, C. A. An evaluation of the continuous progress concept of instruction with university students. Paper presented at the annual meeting of the American Educational Research Association, Chicago, Illinois, 1968.

Morgan, G. A. V. Review of the Iowa Tests of Basic Skills. Buros, O. K. (Ed.) , *The Fifth Mental Measurements Yearbook*. New Jersey: The Gryphon Press, 1959, pp. 34-36.

Morrison, H. C. *The practice of teaching in the secondary schools*. Chicago: University of Chicago Press, 1926.

Myers, J. L. Fundamentals of experimental design. Boston: Allyn & Bacon, 1966.

Pelletti, J. C. The effects of graphic roles on learning geography materials in the middle grades. Unpublished Ed.D. dissertation,

Polya, G. *Patterns of plausible inference mathematics and plausible* reasoning. Vol. 1. Princeton: Princeton University Press, 1954.

Popham, W. J. and Husek, T. R. "Implications of criterion - referenced measurement." *Criterion - Referenced Measurement*. Englewood Cliffs, New Jersey: Educational Technology Publications, 1971.

Powers, R. D., Sumner, W. A., & Kearl, B. E. Recalculation of Readability Formulas, *Journal of Educational Psychology*, 49, No. 2, 1958.

Rice, M. J. Variables in mastery learning in elementary social studies Research proposal submitted to the National Institute of Education, 1973.

Romberg, A., Shepler, J., & King, I. *Mastery learning and retention*. Technical Report No. 151, Wisconsin Research and Development Center for Cognitive Learning, The University of Wisconsin, Madison, Wisconson, 1970.

Russell, D. H. *Children learn to read*. New York: Ginn and Co., 1961.

Sears, P. S. Levels of aspiration in academically successful and unsuccessful children. *Journal of Abnormal Social Psychology*, 35, 1940, 498-536.

Seegers, J. C. 'Vocabulary problems of the elementary school,' *Elementary English Review*, Vol. 16, May, 1939.

Skinner, B. F. The science of learning and the art of teaching. *Harvard Educational Review*, 24 (1954), 86-97.

Spache, A. 'The vocabulary tests of the revised Stanford -Binet as independent measures of intelligence; *Journal of Educational Research*, Vol. 36, March, 1943.

Steinbrink, J. E. The effects of advance organizers for teaching geography to disadvantaged rural black elementary students. Unpublished Ed.D. dissertation, University of Georgia, 1970.

Stringer, L. A., & Glidewell, J. C. Early detection of emotional illnesses in school children. Final Report. St. Louis, Missouri: St. Louis Health Department, Division of Research and Development, 1967.

Stuempfig, D. W., & Maehr, M. L. The effects of conceptual structure and personal quality of feedback on motivation. *Child Development*, 1970, 41.

Tatsuoka, M. M. *Multivariate analysis: Techniques for educational and psychological research*. New York: Wiley, 1971.

Thomas, G. The use of programmed instruction for teaching anthropology in the fifth grade. Unpublished Ed.D. dissertation, University of Georgia, 1967.

Thorndike, R. L. and Hagen, E. *Measurement and evaluation in psychology and education*. New York: John Wiley and Sons, 1969.

Tierney, M. L. A comparison of feedback/correction procedures used in mastery learning strategies. Unpublished Master's thesis, University of Califonia, Santa Barbara, 1973.

Torshen, K. The relation of classroom evaluation to students' self-concepts and mental health. Unpublished Ph.D. dissertation, University of Chicago, 1969.

Traxler, A. E. 'The relationship between vocabulary and general achievement in the elementary school; *Elementary School Journal*, Vol. 45, February, 1945.

U. S. Bureau of Statistics. Statistical abstract of the United States: 1971 (92 ed.) U. S. Government Printing Office, Washington, D. C., 1971.

Walker, H. M., & Lev, J. *Elementary statistical methods*. (Rev. Ed.) New York: Holt, Rinehart & Winston, 1958.

Wallen, N. E., & Travers, R. M. W. 'Analysis and investigation of teaching methods', *Handbook of Research on Teaching*. Chicago: Rand McNally, 1963.

Walsh, J. A. An evaluation of pupil performance in the Anthropology Curriculum Project, grades 1, 2, 4, and 5. Unpublished manuscript, University of Georgia, 1967.

Washburne, C. W. Educational measurement as a key to individualizing instruction and promotions. *Journal of Educational Research*, 5 (1922), 195-206.

Washburne, C. W., Vogel, M., & Gray, W. S. Results of practical experiments in fitting schools to individuals. Supplementary educational monograph, *Journal of Educational Research*, Bloomington, Illinois: Public School Publishing, 1926.

Webster's Third New International Dictionary. Springfield, Mass.: G. and C. Merriam Co., 1971.

Wentling, T. L. Mastery versus non-mastery instruction with varying test item feedback instruments. Journal of Educational Psychology (in press), 1972.

APPENDIX A

Student Text: Functions of Cities
used by both Mastery and Non-
Mastery Treatment Groups

A complete set of the unit Functions of Cities may be ordered from the Geography Curriculum Project, 107 Dudley Hall, University of Georgia, Athens, Georgia 30602.

APPENDIX B

Student Workbook for the Non-Mastery Treatment Group

APPENDIX C

Student Workbook for the Mastery Treatment Group

APPENDIX D

List of Major Facts and Concepts to be Learned

List of Major Facts and Concepts to be Learned

I. Facts: The facts are too numerous to mention here - See Appendix A (STUDENT TEXT), which incorporates the facts to be learned.

II. Concepts: The majority of the concepts to be learned were listed and briefly defined or described in the glossary. The glossary can be found at the end of the Student Text, pp. 10.1-10.6.

APPENDIX E

Table of Specifications for Achievement Tests

Table of Specifications: 40 Item Multiple Choice Achievement Test

Chapter Content Blocks	Knowledge	Application or Transfer	Total
1. Economic base and function	1,2,29, 7.5%	3 2.5%	10%
2. Durban: Port City	4,5 5%	6,8 5%	10%
3. Frankfurt: Commercial City	10,11,12,13 10%	14 2.5%	12.5%
4. Pittsburgh: Industrial City	7,15,16,17,18 12.5%		12.5%
5. Brasilia: Government City	20,22,23 7.5%	21,24 5%	12.5%
6. Surfers Paradise: Resort City	26,27,28,31 10%	30,33	15%
7. Benares: Religious City	32,34,35 7.5%		7.5%
8. Mexico City: Dominant City	37,38 5%		5%
9. Tokyo: Super City	39,40 5%		5%
General Terms	9,25,36 7.5%	19 2.5%	10%
Total	77.5%	22.5%	100%

The 24 item recall test items were all knowledge items based on the content available in each chapter of the unit Functions of Cities.

See Appendix F for a copy of the recall test.

APPENDIX F

The 40 Item Multiple Choice and the 24 Item Recall Geography Achievement Test

APPENDIX G

'Report from Teachers' Weekly Report Form

APPENDIX H

The following forms were used:

1. Directions to Teachers: Non-Mastery and Mastery Instructions.
2. Teacher Information Sheet.
3. School Characteristics Information Sheet.